

## Tuần 4 + 5: Định danh vi sinh vật và xác định sự có mặt của gen chức năng ở vi sinh vật

### Tuần 4: Xử lý trình tự gen mã hóa 16S rRNA và định danh vi sinh vật

Vi sinh vật trong môi trường có nhiều tiềm năng ứng dụng nhờ khả năng trao đổi chất đa dạng, có khả năng sinh trưởng và phát triển với tốc độ lớn, nhu cầu dinh dưỡng đơn giản, có thể sử dụng nguồn nguyên liệu, phế phụ phẩm rẻ tiền từ công nghiệp và nông nghiệp. Việc khai thác vi sinh vật trong môi trường cho các ứng dụng sản xuất enzyme, các sản phẩm có hoạt tính sinh học hoặc trong xử lý chất thải trong môi trường đều bắt đầu từ phân lập chúng từ môi trường. Sau đó các đặc điểm sinh hóa và phân loại chúng được thực hiện trước khi có thể khai thác chúng.

*Trong các bài thực hành ở hai tuần cuối này, bạn sẽ định danh vi khuẩn bằng so sánh trình tự đoạn gen mã hóa 16S rRNA với các trình tự gen mã hóa 16S rRNA trong cơ sở dữ liệu NCBI, và thiết kế mồi cho phản ứng PCR phát hiện sự có mặt của một đoạn gen quy định tính trạng quan tâm của chủng vi khuẩn đó.*

1. Bạn phân lập được một chủng vi khuẩn và định danh nó bằng cách khuếch đại đoạn gen mã hóa 16S rRNA (kích thước khoảng 1500 bp) và xác định trình tự của nó bằng phương pháp Sanger sequencing hai chiều. Kết quả của giải trình tự là hai trace file thể hiện trình tự được xác định chiều xuôi sử dụng mồi 27F và chiều ngược sử dụng mồi 1492R. Hai file được đặt tên **tên\_chủng\_27F.ab1** và **tên\_chủng\_1492R.ab1** có trên Teams của nhóm thí nghiệm.

a) Từ hai trình tự này, tạo trình tự consensus của đoạn gen mã hóa 16S rRNA của chủng.

Hướng dẫn cách tạo trình tự consensus bằng MEGA12 và Bioedit

- Mở phần mềm MEGA12, nhấn trái chuột vào **Align**, chọn **Edit/View Sequencer File (Trace)**, chọn file trình tự từ đầu 27F, cửa sổ **TraceEditor** hiện ra, bạn kiểm tra cẩn thận các tín hiệu đọc, mức độ nhiễu, chọn 25 – 30 nucleotide đầu tiên và chọn từ nucleotide thứ 850 trở đi và xóa các đoạn được chọn bằng cách ấn nút Delete trên bàn phím. Trên cửa sổ **TraceEditor**, chọn **File/Export FASTA** file để lưu file fasta của trình tự từ đầu 27F
- Mở trình tự từ đầu 1492R tương tự như bước trên, sau khi đã xóa các vùng trình tự nhiễu, chọn **Edit/Reverse Complement**, sau đó lưu file fasta tương tự như bước trên.
- Mở **Bioedit**, chọn **File/New Alignment**, cửa sổ alignment hiện ra. Sau đó chọn **File/Import/Sequence alignment file**, chọn 2 trình tự fasta được tạo ra ở hai bước trên.
- So sánh hai trình tự với nhau theo các bước sau: Chọn **Edit/Select all sequences** (hoặc Ctrl+A), khi đó hai trình tự sẽ được highlighted. Chọn **Sequence/Pairwise alignment/Align two sequences (allow ends to slide)**, sau khi thực hiện

alignment, có hai cửa sổ mới xuất hiện: cửa sổ output có các trình tự được align với nhau và cửa sổ tổng kết các thông số của output.

- Tạo trình tự consensus: Ấn chuột vào cửa sổ output ở bước trên, sau đó chọn **Accessory Application/CAP contig assembly program**, cửa sổ CAP contig assembly program interface mở ra, ấn **Run Application**, sau đó Windows Terminal hiện ra thông báo các trạng thái của chương trình CAP contig assembly (initiate, make, pair, assemble, repair, form tree và show kết quả, con chuột nháy ở dòng cuối của Terminal, ấn **enter**, cửa sổ output của assembly xuất hiện, trong đó trình tự consensus là trình tự thứ ba.
- Lưu trình tự consensus gen mã hóa 16S rRNA: ấn chuột vào cửa sổ output của contig assembly, chọn trình tự consensus này, sau đó chọn **Edit/copy sequences to clipboard (Fasta format)**, mở app Notepad, paste trình tự và lưu file.

**Copy trình tự dạng fasta xuống ô trống dưới đây**

>Contig-0

```
GAACGCACCGCAAGGTGAGTGGCAGACGGGTGAGTAACACGTGGGGATCTGCCCAATGGTACGGAATAGCTCCG
GGAAACTGGAATTAATACCGTATGAGCCCGCAAGGGGAAAGATTTATCGCCATTGGATGAACCCGCGTCGGATT
AGCTAGTTGGTGTGGTAAAGGCGCACCAAGGCGACGATCCGTAGCTGGTCTGAGAGGATGATCAGCCACACTGG
GACTGAGACACGGCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCA
GCCATGCCGCGTGAGTGATGAAGGCCTTAGGGTTGTAAAGCTCTTTCGCCGACGAAGATAATGACGGTAGTCGG
AGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGTAATACGAAGGGGGCTAGCGTTGTTCGGAATCACTG
GGCGTAAAGCGCACGTAGGCGGATTGTTAAGTCAGGGGTGAAAGCCTGGAGCTCAACTCCAGAACTGCCCTTGA
TACTGGCAATCTCGAGTCCGGAAGAGGTAAGTGGAAGTGCAGAGTGTAAGGTTGAAATTCGTAGATATTCGCAAG
AACACCAGTGGCGAAGGCGGCTTACTGGTCCGGTACTGACGCTGAGGTGCGAAAGCGTGGGGAGCAAACAGGAT
TAGATACCCTGGTAGTCCACGCCGTAAACGATGGAGGCTAGCCGTTGGCGAGCATGCTCGTCAGTGGCGCAGCT
AACGCATTAAGCCTCCCGCCTGGGGAGTACGGTCGCAAGATTAAGAACTCAAAGGAATTGACGGGGGCCGCACA
AGCGGTGGAGCATGTGGTTTAATTGCAAGCAACGCGCAGAACCTTACCAGCCTTTGACATGCCCCGGAATTGGA
TCAGAGATGAACCAAGCTCTTCGGAGCCGGGGACACAGGTGCTGCATGGCTGTCGTCAGCTCGTGTCGTGAGAT
GTTGGGTAAAGTCCCGCAACGAGCGCAACCTCGCCCTTAGTTGCCATCATTCAGTTGGGCACTCTAGGGGGAC
TGCCGGTGATAAGCCGAGAGGAAGGTGGGGATGACGTCAAGTCCTCATGGCCCTTACGGGCTGGGCTACACACG
TGCTACAATGGCGGTGACAGTGGGACGCGAACCCGCGAGGGTGAGCAAATCTCCAAAAGCCGTCTCAGTTCGGA
TTGCACTCTGCAACTCGAGTGCATGAAGTTGGAATCGCTAGTAATCGTGGATCAGCACGCCACGGTGAATACGT
TCCCGGGCCTTGTACACACCGCCCGTCACACCATGGGAGTTGGCTTTACCCGAAGGCGCTGCGCTAACCCGCAA
```

**2. Sử dụng công cụ blastn của NCBI để xác định danh tính của chủng vi khuẩn này, với các thông tin sau:**

+ Cơ sở dữ liệu để so sánh: rRNA/ITS databases/16S ribosomal RNA sequences (bacteria and archaea)

+ Exclude: chọn Uncultured/environmental sample sequences

a) Lập bảng thống kê sự tương đồng về trình tự và kết luận về danh tính của chủng với 5 trình tự đầu tiên khi sắp xếp kết quả tăng dần giá trị E-value

Tên loài/chủng	Tỷ lệ tương đồng (%)	Mã hiệu trình tự
<i>Ancylobacter novellus</i> DSM 506	99.25	NR_074219.1
<i>Ancylobacter koreensis</i>	99.10	NR_113962.1
<i>Ancylobacter koreensis</i>	99.10	NR_041013.1
<i>Ancylobacter novellus</i> DSM 506	98.50	NR_025859.1
<i>Ancylobacter lacus</i>	97.97	NR_180559.1

Kết luận:

b) Sử dụng cơ sở dữ liệu Taxonomy trên NCBI, **bạn hãy cho biết vị trí phân loại của chủng nghiên cứu**

Hyphomicrobiales / Xanthobacteraceae / Ancylobacter

3. Xây dựng cây phát sinh loài thể hiện mối quan hệ tiến hóa của chủng với các chủng có quan hệ gần gũi.

Thu thập các trình tự mã hóa 16S rRNA (ít nhất 10 trình tự) của các chủng có quan hệ loài gần gũi (trong cùng một họ - **family**) và của một chủng không thuộc ngành của chủng nghiên cứu để làm **outgroup** cho cây phát sinh loài.

a) Tải trình tự fasta và lưu vào các file fasta riêng biệt hoặc lưu vào cùng một file.

Copy các trình tự dạng fasta vào ô trống dưới đây

(\*các trình tự này là bản đã được alignment bằng MEGA):

```
#mega
!Title all_seq_final;
!Format DataType=DNA indel=- CodeTable=Standard;

!Domain=Data property=Coding CodonStart=1;
#Contig-0
GAACGCA--CCGCAAG-----GTGAGTGGCAGACGGGTGAGTAACACGTGGG
GATCTGCCCAATGGTACGGAATAGCTCCGGGAAACTGGAATTAATACCGTATGAGCCCGC
AAGGGGAAAGA-----TTTATCGCCATTGGATGAACCCGCGTCGGATTAGC
```

TAGTTGGTGTGG-TAAAGGCGCACCAAGGCGACGATCCGTAGCTGGTCTGAGAGGATGAT  
CAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATAT  
TGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCTTAGGGTT  
GTAAAGCTCTTTGCGCGACGAAGA-----TAATGACGGTA  
GTCGAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGTAATACGAAGGGGGCTA  
GCGTTGTTCGGAATCACTGGGCGTAAAGCG--CACGTAGGCGGATTGTTAAGTCAGGGGT  
GAAAGCCTGGAGCTCAACTCCAGAACTGCCCTTGATACTGGCAATCTCGAGTCCGGAAGA  
GGTAAGTGGAAGTGCAGAGTGTAGAGGTGAAATTCGTAGATATTCGCAAGAACACCAGTGG  
CGAAGGCGGCTTACTGGTCCGGTACTGACGCTGAGGTGCGAAAGCGTGCGGAGCAAACAG  
GATTAGATACCCTGGTAGTCCACGCGTAAACGATGGAGGCTAGCC-GTTGGCGAGCATG  
CTCGTCAGTGGCGCAGCTAACGCATTAAGCCTCCCGCCTGGGGAGTACGGTCGCAAGATT  
AAAACCTAAAGGAATTGACGGGGGGCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGAA  
GCAACGCGCAGAACCTTACCAGCCTTTGACATGCCCCGGAATTGGATCAGAGATGAACCA  
AGCTCTT-CGG-AGCCGGGGACACAGGTGCTGCATGGCT-GTCGTCAGCTCGTGTCTGA  
GATGTTGGGTAAAGTCCCGCAACGAGCGCAACCCTCGCCCTTAGTTGCCATCATTAGTT  
GGGCACTCTAGGGGGACTGCCGGTGATAAGCCGAGAGGAAGGTGGGGATGACGTCAAGTC  
CTCATGGCCCTTACGGGCTGGGCTACACACGTGCTACAATGGCGGTGACAGTGGGACGCG  
AACCCGCGAGGGTGAGCAAATCTCCAAAAGC-CGTCTCAGTTCGGATTGCACTCTGCAAC  
TCGAGTGCATGAAGTTGGAATCGTAGTAATCGTGGATCAGCACGCCACGGTGAATACGT  
TCCCGGGCCTTGTACACACCGCCCGTCACACCATGGGAGTTGGCTTTACCCGAAGGCGCT  
GCGCTAACCC-GCAA

#NR\_074219.1\_Ancylobacter\_novellus\_DSM\_506\_16S\_ribosomal\_RNA\_partial\_sequence

GAACGCA--CCGCAAG-----GTGAGTGGCAGACGGGTGAGTAACACGTGGG  
GATCTGCCCCAATGGTACGGAATAGCTCCGGGAAACTGGAATTAATACCGTATGTGCCCGC  
AAGGGGAAAGA-----TTTATCGCCATTGGATGAACCCGCGTCGGATTAGC  
TAGTTGGTGTGG-TAAAGGCGCACCAAGGCGACGATCCGTAGCTGGTCTGAGAGGATGAT  
CAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATAT  
TGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCTTAGGGTT  
GTAAAGCTCTTTGCGCGACGAAGA-----TAATGACGGTA  
GTCGAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGTAATACGAAGGGGGCTA  
GCGTTGTTCGGAATCACTGGGCGTAAAGCG--CACGTAGGCGGATTGTTAAGTCAGGGGT  
GAAAGCCTGGAGCTCAACTCCAGAACTGCCCTTGATACTGGCAATCTCGAGTCCGGAAGA

GGTAAGTGGAAGTGCAGTGTAGAGGTGAAATTCGTAGATATTCGCAAGAACACCAGTGG  
CGAAGGCGGCTTACTGGTCCGGTACTGACGCTGAGGTGCGAAAGCGTGGGGAGCAAACAG  
GATTAGATACCCTGGTAGTCCACGCCGTAAACGATGGAGGCTAGCC-GTTGGTGAGCATG  
CTCATCAGTGGCGCAGCTAACGCATTAAGCCTCCCGCCTGGGGAGTACGGTCGCAAGATT  
AAAAC TCAAAGGAATTGACGGGGGGCCCGCACAAAGCGGTGGAGCATGTGGTTTAATTCGAA  
GCAACGCGCAGAACCTTACCAGCCTTTGACATGTCTCGGAATTGGACCAGAGATGGACCA  
AGCTCTT-CGG-AGCCGGGAACACAGGTGCTGTCATGGCT-GTCGTCAGCTCGTGTCTGA  
GATGTTGGGTAAAGTCCCGCAACGAGCGCAACCCTCGCCCTTAGTTGCCATCATTAAAGTT  
GGGCACTCTAGGGGGACTGCCGGTGATAAGCCGAGAGGAAGGTGGGGATGACGTCAAGTC  
CTCATGGCCCTTACGGGCTGGGCTACACACGTGCTACAATGGCGGTGACAGTGGGATGCG  
AACCCGCGAGGGTGAGCAAATCTCCAAAAGC-CGTCTCAGTTCGGATTGCACTCTGCAAC  
TCGAGTGCATGAAGTTGGAATCGCTAGTAATCGTGGATCAGCACGCCACGGTGAATACGT  
TCCCGGGCCTTGTACACACCGCCCGTCACACCATGGGAGTTGGCTTTACCCGAAGGCGCT  
GCGCTAACCC-GCAA

#NR\_113962.1\_Ancylobacter\_koreensis\_strain\_NBRC\_100963\_16S\_ribosomal\_RNA\_partial\_seque  
nce

GAACGCA--CCGCAAG-----GTGAGTGGCAGACGGGTGAGTAACACGTGGG  
GATCTGCCCAATGGTACGGAATAGCTCCGGGAAACTGGAATTAATACCGTATGTGCCCTT  
CGGGGGAAAGA-----TTTATCGCCATTGGATGAACCCGCGTCGGATTAGC  
TAGTTGGTGTGG-TAAAGGCGCACCAAGGCGACGATCCGTAGCTGGTCTGAGAGGATGAT  
CAGCCACACTGGGACTGAGACACGGCCAGACTCCTACGGGAGGCAGCAGTGGGGAATAT  
TGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCTTAGGGTT  
GTAAAGCTCTTTTCGCCGACGAAGA-----TAATGACGGTA  
GTCGGAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGTAATACGAAGGGGGCTA  
GCGTTGTTTCGGAATCACTGGGCGTAAAGCG--CACGTAGGCGGATTGTTAAGTCAGGGGT  
GAAAGCCTGGAGCTCAACTCCAGAAGTGCCTTGATACTGGCAATCTCGAGTCCGGAAGA  
GGTAAGTGGAAGTGCAGTGTAGAGGTGAAATTCGTAGATATTCGCAAGAACACCAGTGG  
CGAAGGCGGCTTACTGGTCCGGTACTGACGCTGAGGTGCGAAAGCGTGGGGAGCAAACAG  
GATTAGATACCCTGGTAGTCCACGCCGTAAACGATGGAGGCTAGCC-GTTGGTGAGCATG  
CTCATCAGTGGCGCAGCTAACGCATTAAGCCTCCCGCCTGGGGAGTACGGTCGCAAGATT  
AAAAC TCAAAGGAATTGACGGGGGGCCCGCACAAAGCGGTGGAGCATGTGGTTTAATTCGAA  
GCAACGCGCAGAACCTTACCAGCCTTTGACATGTCCCGGAATTGGATCAGAGATGGACCA

AGCTCTT-CGG-AGCCGGAACACAGGTGCTGCATGGCT-GTCGTCAGCTCGTGTCTGA  
GATGTTGGGTAAAGTCCCGCAACGAGCGCAACCCTCGCCCTTAGTTGCCATCATTAGTT  
GGGCACTCTAGGGGACTGCCGGTGATAAGCCGAGAGGAAGGTGGGGATGACGTCAAGTC  
CTCATGGCCCTTACGGGCTGGGCTACACACGTGCTACAATGGCGGTGACAGTGGGAAGCG  
AACCCGCGAGGGTGAGCAAATCTCCAAAAGC-CGTCTCAGTTCGGATTGCACTCTGCAAC  
TCGAGTGCATGAAGTTGGAATCGCTAGTAATCGTGGATCAGCACGCCACGGTGAATACGT  
TCCCGGGCCTTGTTACACACCGCCCGTCACACCATGGGAGTTGGTTTTACCCGAAGGCGCT  
GCGCTAACCC-GCAA

#NR\_180560.1\_Ancylobacter\_plantiphilus\_strain\_1TC\_16S\_ribosomal\_RNA\_partial\_sequence

GAACGCC--CCGCAAG-----GGGAGTGGCAGACGGGTGAGTAACACGTGGG  
GATCTGCCCCAATGGTACGGAATAGCTCCGGGAACTGGAATTAATACCGTATAAGCCCGC  
AAGGGGAAAGA-----TTTATCGCCATTGGATGAACCCGCGTCGGATTAGC  
TAGTTGGTGGG-TAAAGGCCTACCAAGGCGACGATCCGTAGCTGGTCTGAGAGGATGAT  
CAGCCACACTGGGACTGAGACACGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATAT  
TGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCTTAGGGTT  
GTAAAGCTCTTTCCGCGACGAAGA-----TAATGACGGTA  
GTCGGAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGTAATACGAAGGGGGCTA  
GCGTTGTTCCGAATCACTGGGCGTAAAGCG--CACGTAGGCGGATTGTTAAGTCAGAGGT  
GAAAGCCTGGAGCTCAACTCCAGAACTGCCCTTGATACTGGCAATCTCGAGTCCGGAAGA  
GGTAAGTGGAAGTGCAGAGTGTAGAGGTGAAATTCGTAGATATTCGCAAGAACACCAGTGG  
CGAAGGCGGCTTACTGGTCCGGTACTGACGCTGAGGTGCGAAAGCGTGGGGAGCAAACAG  
GATTAGATACCCTGGTAGTCCACGCCGTAAACGATGGAGGCTAGCC-GTTGGTGAGCATG  
CTCATCAGTGGCGCAGCTAACGCATTAAGCCTCCCGCTGGGGAGTACGGTCGCAAGATT  
AAAAC TCAAAGGAATTGACGGGGGCCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGAA  
GCAACGCGCAGAACCTTACCAGCCTTTGACATGCCCAGGACAGTTACCAGAGATGGTTTC  
CTCTCTT-CGG-AGCCTGGGACACAGGTGCTGCATGGCT-GTCGTCAGCTCGTGTCTGA  
GATGTTGGGTAAAGTCCCGCAACGAGCGCAACCCTCGCCCTTAGTTGCCATCATTAGTT  
GGGCACTCTAGGGGACTGCCGGTGATAAGCCGAGAGGAAGGTGGGGATGACGTCAAGTC  
CTCATGGCCCTTACGGGCTGGGCTACACACGTGCTACAATGGCGGTGACAATGGGATGCA  
AGACCGCGAGGTTGAGCAAATCTCCAAAAGC-CGTCTCAGTTCGGATTGCACTCTGCAAC  
TCGAGTGCATGAAGTTGGAATCGCTAGTAATCGTGGATCAGCACGCCACGGTGAATACGT  
TCCCGGGCCTTGTTACACACCGCCCGTCACACCATGGGAGTTGGTTTTACCCGAAGGCGTT

GCGCTAACCC-GCAA

#NR\_042794.1\_Ancylobacter\_vacuolatus\_strain\_DSM\_1277\_16S\_ribosomal\_RNA\_partial\_sequence

GAACGCC--CCGCAAG-----GGGAGTGGCAGACGGGTGAGTAACACGTGGG  
GATCTGCCCCAATGGTACGGAATAATTTCTGGGAAACTGGGACTAATACCGTATGAGCCCGC  
AAGGGGAAAGA-----TTTATCGCCATTGGATGAACCCGCGTCGGATTAGC  
TAGTTGGTGTGG-TAAAGGCGCACCAAGGCGACGATCCGTAGCTGGTCTGAGAGGATGAT  
CAGCCACACTGGGACTGAGACACGGCCCA?ACTCCTACGGGAGGCAGCAGTGGGGAATAT  
TGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCTTAGGGTT  
GTAAAGCTCTTTCTGCCGACGAAGA-----TAATGACGGTA  
GTCGGAGAAAAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGTAATACGAAGGGGGCTA  
GCGTTGTTCCGAATCACTGGGCGTAAAGCG--CACGTAGGCGGATATTTAAGTCAGGGGT  
GAAAGCCTGGAGCTCAACTCCAGAACTGCCCTTGATACTGGGTATCTCGAGTCCGGAAGA  
GGTAAGTGGAAC TGCGAGTGTAGAGGTGAAATTCGTAGATATTCGCAAGAACACCAGTGG  
CGAAGGCGGCTTACTGGTCCGGTACTGACGCTGAGGTGCGAAAGCGTGGGGAGCAAACAG  
GATTAGATACCCTGGTAGTCCACGCCGTAAACGATGGAGGCTAGCC-GTTGGTGAGCATG  
CTCATCAGTGGCGCAGCTAACGCATTAAGCCTCCCGCCTGGGGAGTACGGTCGCAAGATT  
AAAAC TCAAAGGAATTGACGGGGGGCCG CACAAGCGGTGGAGCATGTGGTTTAATTCGAA  
GCAACGCGCAGAACCTTACCAGCCTTTGACATGTCCCGGACGGTTACCAGAGATGGTTTC  
TTCTCTT-CGG-AGCCGGGAACACAGGTGCTGCATGGCT-GTCGTGAGCTCGTGTCTGA  
GATGTTGGGTAAAGTCCCGCAACGAGCGCAACCCTCGCCCTTAGTTGCCATCATTAAGTT  
GGGCACTCTAGGGGGACTGCCGGTGATAAGCCGAGAGGAAGGTGGGGATGACGTCAAGTC  
CTCATGGCCCTTACGGGTGGGCTACACACGTGCTACAATGGCGGTGACAGTGGGAAGCG  
AACCTGCGAGGGTAAGCAAATCTCCAAAAGC-CGTCTCAGTTCGGATTGCAC TCTGCAAC  
TCGAGTGCATGAAGTTGGAATCGCTAGTAATCGTGGATCAGCATGCCACGGTGAATACGT  
TCCCGGGCCTTGTTACACACCGCCCGTCACACCATGGGAGTTGGTTTTACCCGAAGGCGCT  
GCGCTAACCC-GCAA

#NR\_044737.2\_Ancylobacter\_aquaticus\_strain\_ATCC\_25396\_16S\_ribosomal\_RNA\_partial\_sequence

GAACGCC--CCGCAAG-----GGGGTGGCAGACGGGTGAGTAACACGTGGG  
GATCTGCC?AATGGTACGGAATAATTTCTGGGAAACTGGGACT?ATACCGTATGAGCCCGC  
AAGGGGAAAGA-----TTTATCGCCATTGGATGAACCCGCGTCGGATTAGC

TAGTTGGTGTGG-TAAAGGCGCACCAAGGCGACGATCCGTAGCT?GTCTGAGAGGATGAT  
CAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATAT  
TGGACAATGGGCGCAAGCC??ATCCAGCCATGCCGCGTGAGTGATGAAGGCCTTAGGGTT  
GTAAAGCTCTTTGCCGACGAAGA-----TAATGACGGTA  
GTCGAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGTAATACGAAGGGGGC?A  
GCGTTGTTCGGAATCACTGGGCGTAAAGCG--CACGTAGGCGGATATTTAAGTCAGGGGT  
GAAAGCCTGGAGCTCAACTCCAGAACTGCCCTTGATACTGGGTATCTCGAGTCCGGAAGA  
GGTAAGTGGAAGTGCAGGTGTAGAGGTGAAATTCGTAGATATTCGCAAGAACACCAGTGG  
CGAAGGCGGCTTACTGGTCCGGTACTGACGCTGAGGTGC?AAAGCGTGGGAGCAAACAG  
GATTAGATACCCTGGTAGTCCACGCGTAAACGATGGAGGCTAGCC-GTTGGTGAGCATG  
CTCATCAGTGGCGCAGCTAACGCATTAAGCCTCCCGCCTGGGGAGTACGGTCGCAAGATT  
AAAACCTAAAGGAATTGACGGGGGCC?GCACAAGCGGTGGAGCATGTGGTTTAATTCGAA  
??AACGCGCAGAACCTTACCAGCCTTTGACATGTCCCGGACGGTTACCAGAGATGGTTTC  
TTCTCTT-CGG-AGCCGGGAACACAGGTGCTGCATGGCT-GTCGTCAGCTCGTGTCTGA  
GATGTTGGGTAAAGTCCCGCAACGAGCGCAACCCTCGCCCTTAGTTGCCATCATTAAAGTT  
GGGCACTCTAGGGGGACTGCCGGTGATAAGCCGAGAGGAAGGTGGGGATGACGTCAAGTC  
CTCATGGCCCTTACGGGCTGGGCTACACACGTGCTACAATGGCGGTGACAGTGGGAAGCG  
AACCTGCGAGGGTAAGCAAATCTCCAAAAGC-CGTCTCAGTTCGGATTGCACTCTGCAAC  
TCGAGTGCATGAAGTTGGAATCGTAGTAATCGTGGATCAGCATGCCACGGTGAATACGT  
TCCCGGGC?TTGTACACACCGCCCGTCACACCATGGGAGTT?GTTTTACCGAAGGCGCT  
GCGCTAACCC-GCAA

#NR\_025172.1\_Xanthobacter\_aminoxidans\_strain\_14a\_16S\_ribosomal\_RNA\_partial\_sequence

GAGCGCC--CAGCAAT-----GGGAGCGGCAGACGGGTGAGTAACACGTGGG  
GATCTACCCAATGGTACGGAATAACCCAGGGAAACTTGGATTAATACCGTATGTGCCCTT  
CGGGGAAAGA-----TTTATCGCCATTGGATGAACCCGCGTCGGATTAGC  
TAGTTGGTGAGG-TAAAGGCTCACCAAGGCGACGATCCGTAGCTGGTCTGAGAGGATGAT  
CAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATAT  
TGGACAATCGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCTTAGGGTT  
GTAAAGCTCTTTGCCGGTGAAGA-----TAATGACGGTA  
ACCGGAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGTAATACGAAGGGGGCTA  
GCGTTGCTCGGAATCACTGGGCGTAAAGCG--CACGTAGGCGGATCGTTAAGTCAGGGGT  
GAAATCCTGGAGCTCAACTCCAGAACTGCCCTTGATACTGGCGATCTCGAGTTCGAGAGA

GGTTGGTGGAACCTCCGAGTGTAGAGGTGAAATTCGTAGATATTCGGAAGAACACCAGTGG  
CGAAGGCGGCCAACTGGCTCGATACTGACGCTGAGGTGCGAAAGCGTGGGGAGCAAACAG  
GATTAGATACCCTGGTAGTCCACGCCGTAAACGATGGATGCTAGCCTGTGGGGGGTTTA  
CCTCTCAGTGGCGCAGCTAACGTCTTAAGCATCCCGCCTGGGGAGTACGGTCGCAAGATT  
AAAAC TCAAAGGAATTGACGGGGGCCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGAA  
GCAACGCGCAGAACCTTACCAGCCTTTGACATGGCAGG-ACGACTTCCGGAGACGGATTT  
CTTCCAG-CAATGGACCTCGACACAGGTGCTGCATGGCT-GTCGTCAGCTCGTGTCTGA  
GATGTTGGGTAAAGTCCCGCAACGAGCGCAACCCTCGCCTCTAGTTGCCATCATTCAAGTT  
GGGCACTCTAGAGGGACTGCCGGTGATAAGCCGAGAGGAAGGTGGGGATGACGTCAAGTT  
CTCATGGCCCTTACGGGCTGGGCTACACACGTGCTACAATGGCGGTGACAGTGGGATGCG  
AACCCGCGAGGGTAAGCAAATCTCCAAAAGC-CGTCTCAGTTCGGATTGCACTCTGCAAC  
TCGAGTGCATGAAGTTGGAATCGCTAGTAATCGTGGATCAGCATGCCACGGTGAATACGT  
TCCCGGGCCTTGTACACACCGCCCGTCACACCATGGGAGTTGGCTTTACCCGAAGGCGCT  
GCGCTAACCC-GCAA

#NR\_179776.1\_Aquabacter\_cavernae\_strain\_Sn-9\_2\_16S\_ribosomal\_RNA\_partial\_sequence

GAGCGCC--CAGCAAT-----GGGAGCGGCAGACGGGTGAGTAACGCGTGGG  
GATCTGCCCAGTGGTACGGAATAATTCCGGGAACTGGGACTAATACCGTATGTGCCCCG  
AAGGGGAAAGA-----TTTATCGCCATCGGATGAACCCGCGTCGGATTAGC  
TAGTTGGTGTGG-TAAAGGCGCACCAAGGCGACGATCCGTAGCTGGTCTGAGAGGATGAT  
CAGCCACACTGGGACTGAGACACGGCCAGACTCCTACGGGAGGCAGCAGTGGGGAATAT  
TGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCTTAGGGTT  
GTAAAGCTCTTTCCGCCGTGAAGA-----TAATGACGGTA  
ACCGGAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGTAATACGAAGGGGGCTA  
GCGTTGCTCGGAATCACTGGGCGTAAAGCG--CACGTAGGCGGGTCGTAAAGTCAGAGGT  
GAAAGCCTGGAGCTCAACTCCAGAACTGCCTTTGATACTGGCGACCTAGAGTTCGAGAGA  
GGTTGGTGGAACCTCCGAGTGTAGAGGTGAAATTCGTAGATATTCGGAAGAACACCAGTGG  
CGAAGGCGGCCAACTGGCTCGATACTGACGCTGAGGTGCGAAAGCGTGGGGAGCAAACAG  
GATTAGATACCCTGGTAGTCCACGCCGTAAACGATGGATGCTAGCC-GTTGGGGAGCTTG  
CTCTTCAGTGGCGCAGCTAACGCCTTAAGCATCCCGCCTGGGGAGTACGGTCGCAAGATT  
AAAAC TCAAAGGAATTGACGGGGGCCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGAA  
GCAACGCGCAGAACCTTACCAGCCTTTGACATGGCAGG-ACGATTTCAGAGATGGATCT  
CTTCCAG-CAATGGACCTGCACACAGGTGCTGCATGGCT-GTCGTCAGCTCGTGTCTGA

GATGTTGGGTAAAGTCCCGCAACGAGCGCAACCCTCGCCTTTAGTTGCCATCATTCAGTT  
GGGCACTCTAAAGGGACTGCCGGTGATAAGCCGCGAGGAAGGTGGGGATGACGTCAAGTC  
CTCATGGCCCTTACGGGCTGGGCTACACACGTGCTACAATGGTGGTGACAGTGGGATGCG  
AAAGGGCGACCTCTAGCAAATCTCCAAAAGC-CATCTCAGTTCGGATTGCACTCTGCAAC  
TCGAGTGCATGAAGTTGGAATCGCTAGTAATCGTGGATCAGCATGCCACGGTGAATACGT  
TCCCGGGCCTTGTACACACCGCCCGTCACACCATGGGAGTTGGTTTTACCCGAAGGCGCT  
GCGCTAACC--GCAA

#NR\_041839.1\_Azorhizobium\_doebereineriae\_strain\_BR5401\_16S\_ribosomal\_RNA\_partial\_sequence

GAGCGCC--CAGCAAT-----GGGAGCGGCAGACGGGTGAGTAACGCGTGGG  
GATCTGCCCCAATGGTACGGAATAACCCAGGGAACTTGGATTAATACCGTATGTGCCCTT  
CGGGGGAAAGA-----TTTATCGCCATTGGATGAACCCGCGTCGGATTAGC  
TAGTTGGTGTGG-TAAAGGCGCACCAAGGCGACGATCCGTAGCTGGTCTGAGAGGATGAT  
CAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATAT  
TGGACAATGGGCGCAAGCCTGATTCAGCCATGCCGCGTGTGTGATGAAGGCCTTAGGGTT  
GTAAAGCACTTTCGCCGGTGAAGA-----TAATGACGGTA  
ACCGGAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGTAATACGAAGGGGGCAA  
GCGTTGCTCGGAATCACTGGGCGTAAAGCG--TACGTAGGCGGGTCGTAAAGTCAGGGGT  
GAAATCCTGGAGCTCAACTCCAGAACTGCCCTTGATACTGGCGATCTTGAGTTCGAGAGA  
GGTTGGTGGAAGTGCAGGTGTAGAGGTGAAATTCGTAGATATTCGCAAGAACACCAGTGG  
CGAAGGCGGCCAACTGGCTCGATACTGACGCTGAGGTACGAAAGCGTGGGGAGCAAACAG  
GATTAGATACCCTGGTAGTCCACGCCGTAAACGATGGATGCTAGCC-GTTAGGCAGCTTG  
CTGCTTAGTGGCGCAGCTAACGCATTAAGCATCCCGCCTGGGGAGTACGGTCGCAAGATT  
AAAAC TCAAAGGAATTGACGGGGGCGCCG CACAAGCGGTGGAGCATGTGGTTTAATTCGAA  
GCAACGCGCAGAACCTTACCAGCCTTTGACATGGCAGG-ACGACTTCCGGAGACGGATTT  
CTTCCAG-CAATGGACCTGCACACAGGTGCTGCATGGCT-GTCGTCAGCTCGTGTCTGA  
GATGTTGGGTAAAGTCCCGCAACGAGCGCAACCCTCGCCTTTAGTTGCCATCATTCAGTT  
GGGCACTCTAAAGGGACTGCCGGTGATAAGCCGCGAGGAAGGTGGGGATGACGTCAAGTC  
CTCATGGCCCTTACGGGCTGGGCTACACACGTGCTACAATGGCGGTGACAGTGGGATGCG  
AGCCTGCGAGGGTGAGCAAATCTCCAAAAGC-CGTCTCAGTTCGGATTGCACTCTGCAAC  
TCGAGTGCATGAAGTTGGAATCGCTAGTAATCGTGGATCAGCATGCCACGGTGAATACGT  
TCCCGGGCCTTGTACACACCGCCCGTCACACCATGGGAGTTGGTTTTACCCGAAGGCGTT

GCGCTAACCTAGCAA

#NR\_115492.1\_Azorhizobium\_caulinodans\_ORF\_571\_16S\_ribosomal\_RNA\_partial\_sequence

GAGCGCC--CAGCAAT-----GGGAGCGGCAGACGGGTGAGTAACGCGTGGG  
GATGTGCCCCAATGGTGCAGGAATAACCCAGGGAACTTGGATTAATACCGCATGTGCCCTT  
CGGGGGAAGA-----TTTATCGCCATTGGATCAACCCGCGTCTGATTAGC  
TAGTTGGTGAGG-TAAAGGCTCACCAAGGCGACGATCAGTAGCTGGTCTGAGAGGATGAT  
CAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATAT  
TGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCTTAGGGTT  
GTAAAGCTCTTTGCGCCGTGAAGA-----TAATGACGGTA  
ACCGGAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGTAATACGA?GGGGCAA  
GC?TTGCTCGGAATCACTGGGCGTAAAGCG--CACGTAGGCGGATCGTTAAGTCAGGGGT  
GAAAGCCTGGAGCTCAACTCCAGAACTGCCCTTGATACTGGCGATCTTGAGTTCGAGAGA  
GGTTGGTGGAATCCGAGT?TAGAGGTGAAATTCGTAGATATTCGGAAGAACCAGT?G  
CGAAGGCGGCCAACTGGCTCGATACTGACGCTGAGGTGCGAAA??GTGGGAGCAAACAG  
GATTAGATACCCTGGTAGTCCACGCCGTAAACGATGGATGCTAGCC-GTTGGGAGCTTG  
CTCTTCAGTGGCGCA??TAACGCCTTAAGCATCCCGCCTGGGGAGTACGGTCGCAAGATT  
AAAACCTCAAAGGAATTGACGGGGGGCCGACACAAGCGGTGGAGCATGTGGTTTAATTCGAA  
GCAACGCGCAGAACCTTACCAGCCTTTGACATGGCAGG-ACGACTTCCGGAGACGGATTT  
CTTCAG-CAATGGACCTGCACACAGGTGCTGCATGGCT-GTCGTGAGCTCGTGTCTGTA  
GATGTTGGGTAAAGTCCCGCAACGAGCGCAACCCTCGCCTTTAGTTGCCATCATTAGTT  
GGGCACTCTAAAGGGACTGCCGGTGATAAGCCGCGAGGAAGGTGGGGATGACGTCAAGTC  
CTCATGGCCCTTACGGGCTGGGCTACACACGTGCTACAATGGCGGTGACAATGGGATGCG  
AGCCT?CGAGGGTGAGCAAATCTCCAAAAGC-CGTCTCAGTTCGGATTGCACTCTGCAAC  
TCGAGT?CATGAAGTTGGAATCGCTAGTAATCGTGGATCAGCATGCCACGGTGAATACGT  
TCCCGGGCCTTGTACACACCG?CCGTACACCATGGGAGTTGGCTTTACCGAAGGCGTT  
GCGCTAACCT--GCAA

#NR\_116004.1\_Labrys\_wisconsinensis\_strain\_W1215\_PCA4\_16S\_ribosomal\_RNA\_partial\_sequence

GAGCGCC--CCGCAAG-----GGGAGCGGCAGACGGGTGAGTAACGCGTGGG  
GATGTACCCGAAGGTACGGAACAACCTCCGGGAACTGGAGCTAATACCGTATGTGCCCCA  
AAGGGGAAGA-----TTTATCGCCTTTGGATCAACCCGCGTCAGATTAGC

TAGTTGGTGAGGGTAATGGCCTACCAAGGCGACGATCTGTAGCTGGTCTGAGAGGATGAT  
CAGCCACACTGGGACTGAGACACGGCCAGACTCCTACGGGAGGCAGCAGTGGGGAATAT  
TGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGACGGCCTTAGGGTT  
GTAAAGCTCTTTTCGACGGGGACGA-----TAATGACGGTA  
CCCGTAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAGGGGGCTA  
GCGTTGTTTCGGAATCACTGGGCGTAAAGCG--CACGTAGGCGGATTGTTAAGTCGGGGGT  
GAAATCCTGAGGCTCAACCTCAGAACTGCCTTCGATACTGGCAATCTCGAGTCCGGAAGA  
GGTTGGTGGAACAGCTAGTGTAGAGGTGAAATTCGTAGATATTAGCTAGAACACCAGTGG  
CGAAGGCGGCCAACTGGTCCGGCACTGACGCTGAGGTGCGAAAGCGTGGGGAGCAAACAG  
GATTAGATACCCTGGTAGTCCACGCCGTAAACGATGGATGCTAGCC-GTTGGGGAGCTTG  
CTCTTCGGTGGCGCAGTCAACACATTAAGCATCCCGCCTGGGGAGTACGGTCGCAAGATT  
AAAACCTAAAGGAATTGACGGGGGGCCCGCACAAAGCGGTGGAGCATGTGGTTTAATTCGAA  
GCAACGCGCAGAACCTTACCAGCCCTTGACATCCCGGTGCGGGCCTTCAGAGACGAAGGC  
CTTCAGTTCGGCTGGACCGGAGACAGGTGCTGCATGGCT-GTCGTCAGCTCGTGTCTGA  
GATGTTGGGTAAAGTCCCGCAACGAGCGCAACCCTCGCCCCTAGTTGCCATCATTAGTT  
GGGCACTCTAGGGGGACTGCCGGTGATAAGCCGCGAGGAAGGTGGGGATGACGTCAAGTC  
CTCATGGCCCTTACGGGCTGGGCTACACACGTGCTACAATGGCGGTGACAGTGGGATGCG  
AAGGGGCGACCCCTTAGCAAATCTCAAAAAGC-CGTCTCAGTTCAGATTGCACTCTGCAAC  
TCGAGTGCATGAAGGTGGAATCGTAGTAATCGTAGATCAGCATGCTACGGTGAATACGT  
TCCCGGGCCTTGTACACACCGCCCGTCACACCATGGGAGTTGGTTTTACCCGAAGGCGCT  
GCGCCAACC--GCAA

#LC777924.1\_Escherichia\_coli\_VANM2\_gene\_for\_16S\_rRNA\_partial\_sequence

GAACGGTAACAGGAAGCACTGCTTTGCTGACGAGTGGCGGACGGGTGAGTAATGTCTGGG  
AAACTGCC?GATGGAGGGGGATAACTACTGGAACGGTAGCTAATACCGCATAACGTGCG  
AAGACCAAAGAGGGGGACCTTAGGGCCTCTTGCCATCGGATGTGCCAGATGGGATTAGC  
TAGTAGGTGGGG-TAACGGCTCACCTAGGCGACGATCCCTAGCTGGTCTGAGAGGATGAC  
CAGCCACACTGGAAGTGAAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATAT  
TGCACAATGGGCGCAAGCCTGATGCAGCCATGCCGCGTGATGAAGAAGGCCTTCGGGTT  
GTAAAGTACTTTTCAGCGGGGAGGAAGGGAGTAAAGTTAATACCTTTGCTCATTGACGTTA  
CCCGCAGAAGAAGCACCGGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAA  
GCGTTAATCGGAATTACTGGGCGTAAAGCGAACACGAGGCGGTTTGTTAAGTCAGATGT  
GAAATCCCCGGGCTCAACCTGGGAAGTGCATCTGATACTGGCAAGCTTGAGTCTCGTAGA

GGGGGG?AGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGATCTGGAGGAATACCGGTGG  
 CGAAGGC--CC--CTGGACGAAGACTGACGCTCAGGTGCGAAAGCGTGGGGAGCAAACAG  
 GATTAGATACCCTGGTAGTCCACGCGTAAACGATGTCGACTTGGAGGTTGTGCCCTTGA  
 GGCGTGGCTTCCGGAGCTAACGCGTTAAGTCGACCGCCTGGGGAGTACGGCCGCAAGGTT  
 AAAACTCAAATGAATTGACGGGGGGCCCGCACAGCGGTGGAGCATGTGGTTTAATTCGAT  
 GCAACGCGAAGAACCTTACCTGGTCTTGACATCCACGG--AAGTTTTCAGAGATGAGAAT  
 GTGCCTT-CGG-GAACCGTGAGACAGGTGCTGCATGGCTTGTCGTGAGCTCGTGTGTGA  
 AATGTTGGGTAAAGTCCCGCAACGAGCGCAACCCTTATCCTTTGTTGCCAGCGGTC--CC  
 GGGAACTCAAAGGAGACTGCCAGTGATAAACTG-GAGGAAGGTGGGGATGACGTCAAGTC  
 ATCATGGCCCTTACGACCAGGGCTACACACGTGCTACAATGGCGCATACAAAGAGAAGCG  
 ACCTCGCGAGAGCAAGCG?ACCTCATAAAGTGCGTCGTAGTCCGGATTGGAGTCTGCAAC  
 T-GACTCCATGAAGTCGGAATCGCTAGTAATCGTGGATCAGAATGCCACGGTGAATACGT  
 TCCCGGGCCTTGTACACACCGCCCGTCACACCATGGGAGTGGGTTGCAAAAGAAGTAGGT  
 AGCTTAACCT-TCGG

**b) Sử dụng phần mềm MEGA12 để thực hiện so sánh nhiều trình tự gen mã hóa 16S rRNA của chủng nghiên cứu và các chủng gần về mặt phân loại. Các bước thực hiện như sau:**

- Mở app **MEGA12**, chọn **Align/Edit/build alignment**, chọn **Create a new alignment → DNA**, cửa sổ Alignment Explorer hiện ra với một trình tự trống mặc định là Sequence 1, chọn trình tự này và ấn nút Delete
- Đưa các trình tự cần so sánh vào từ một hoặc nhiều file đã lưu ở mục 1a và 3a: Chọn **Edit/Insert sequences from file** để đưa các file trình tự fasta vào.
- Align các trình tự với nhau bằng cách: Trên cửa sổ **Alignment Explorer**, chọn **Edit/Select All**, khi đó tất cả các trình tự được nhập vào trước đó sẽ được highlight. Tiếp theo, chọn **Alignment/Align by ClustalW**, cửa sổ thông số của ClustalW hiện ra, giữ nguyên các thông số mặc định và ấn OK. Sau khi chương trình thực hiện xong việc align trình tự, bạn kiểm tra lại alignment, cắt bỏ hai đầu của alignment sao cho số nucleotide của các trình tự trong alignment tương đối giống nhau.
- Xuất dữ liệu alignment: Vẫn trên cửa sổ **Alignment Explorer**, chọn **Data/Export alignment/MEGA format**, lưu file alignment bạn vừa tạo ra ở vị trí và tên phù hợp.

**Copy cửa sổ kết quả so sánh (Alignment explorer) vào ô trống dưới đây.**

```

Pairwise Alignment
Sequence 1: 5292521_C3_27F
Sequence 2: 5292522_C3_1492R
Sequence ends allowed to slide over each other
Alignment score: 8111783

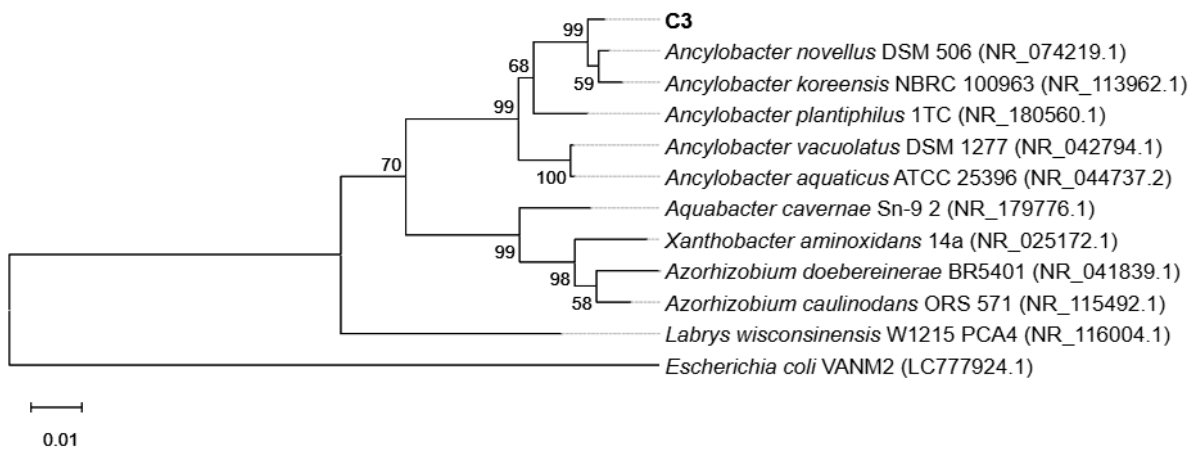
Identities: 0.2747748

```

c) Xây dựng cây phát sinh loài bằng thuật toán Maximum Likelihood hoặc Neighbor-joining với giá trị bootstrap là 1000.

- Tạo cây phát sinh loài: **trên cửa sổ chính** của MEGA12, chọn **Phylogeny/construct/test maximum likelihood tree** hoặc **Phylogeny/construct/test neighbor joining tree**, bạn chọn file chứa các trình tự được align với nhau (MEGA file) để mở file alignment bạn mới tạo ở bước trên. Giữ các thông số Phylogeny reconstruction mặc định, chỉ thay đổi ở **Phylogeny Test: Test of phylogeny** bạn chọn **Standard Bootstrap, Bootstrap replicates**, bạn chọn **1000**.
- Cửa sổ Tree Explorer sẽ xuất hiện sau khi cây phát sinh loài được tạo, bạn thay đổi cây và ấn chuột vào biểu tượng copy to clipboard, paste vào file Powerpoint và tiếp tục chỉnh sửa nếu cần.

Copy cây phát sinh loài vào ô trống dưới đây



## Tuần 5: Thiết kế mồi khuếch đại một đoạn gen chỉ thị ở vi sinh vật

4. **Chủng phân lập** được thể hiện một tính trạng nào đó, tính trạng này đã được xác nhận ở một số **chủng vi khuẩn khác** và các gen quy định tính trạng đó đã được xác định trình tự ở các chủng vi khuẩn này. **Bạn có thể thiết lập một giả thuyết rằng chủng phân lập có mang gen có độ tương đồng cao (ortholog) với các gen đã được nghiên cứu ở các chủng khác. Bạn cần chứng minh giả thuyết này bằng cách xác định sự có mặt của ortholog này ở chủng phân lập bằng phản ứng PCR.**

Để làm được điều đó, bạn cần thiết kế mồi cho phản ứng PCR để khuếch đại đoạn gen quan tâm, việc thiết kế mồi dựa vào các gen ở các chủng đã được nghiên cứu, mồi được thiết kế vào vùng bảo thủ (giống nhau) trên các gen ở các chủng đã được nghiên cứu.

Bạn thực hiện thiết kế mồi theo các bước như sau:

a) Tìm kiếm gen quan tâm ở các chủng vi khuẩn đã được nghiên cứu (5-10 ortholog). Chú ý tìm những trình tự mã hóa toàn bộ (hoặc gần như toàn bộ protein quan tâm, và tìm trong những loài gần nhau trong hệ thống phân loại)

**Copy các trình tự ortholog của gen quan tâm dưới dạng multi-fasta vào ô trống dưới đây**

```
>KX250215.1 Ancylobacter sp. strain TS-1 (aioA) gene, partial cds
GACAACGCGTTTCGGCGTCGACCTCGCCGAGCAGCAGGGGGCGGACACCCCGGCCTGGTACGCCCCGTCCA
TGTACAACATTGTCCGCCAGGACGGCCGAGATGTGCACATCGTCATCAAGCCCGACCGCGAATGCTCGGT
GAACTCCGGCCTCGGATCGGTGCGCGGCGCGCATCGCGGAGATGAGCTATTGCGCCAACGCAACACG
CAGCTCCAGCGCCTTACCGATCCGCTGGTCTGGCGCTACGGCCAGCTGCAACCGACCAGCTGGGACGACG
CGCTCGACCTCGTCGCCCCGGGTGACATCGGCCGTGATCGCCGAACAGGGCGAGGATGGGCTGTTCTGCTC
CGCTTTTCGACCATGGCGGGCGCCGGCGGGCTATGAGAACACCTGGGGCACCGGCAAGCTCTATTTTGGG
GCGATGAAGGTCAAGAACATCCGCATCCACAACCGCCCGGCCTACAATTCCGAGGTTACGGCTCGCGCG
ACATGGGCGTCGGCGAACTGAACAATTGCTATGAGGACGCCAGCTCGCCGACACCATCGTGGCCGTCGG
CACCAACGCGCTGGAAACCCAGACCAACTACTTCCTGAACCATTTGGGTGCCGAATTTGCGCGGCACCTCG
ATGGACAAAAAGAAGGCCGAGTTTCGGCAGCGAGCCGGTGGCTCAGGGCCGGGTGGTCATCGTCCGGTCCGC
GCCGCACAGTGACGGTCAATGCCTGCGAAGTGGAGGCGGGCAAGGACAACGTGCTTCATCTCGCGCTCAA
TTCCGGCACCGACCTCGCGTTGTTCAACGCCTGGTTACCTATGCGGCGGAAAAGGGCTGGATCGACCGG
GAATTCATCGGCGCGTCCACGAAAGACTTCGACAAGGCGGTGCGCGCAACAAGGTCTCGATTGCCGAGG
CGGCGGAGATCACCGGCCTTTCCGAGGCCGACATCGTCAAGGCGATTACCTGGATCGCCGAGCCCCA
```

>KT992350.1 *Rhizobium* sp. strain YE2-4 (aioA) gene, partial cds

CCACTTCTGCATCGTGGGGTGTGGGTACCACGCCTACACGTGGCCGATCAACAAGCAGGGCGGTACCGAT  
CCTCAGAAGAACGTGTTTGGCGTCGATCTGTTCGGAACAACAGCAGGCGGAGACTGACGCCTGGTATTTCGC  
CATCCATGTACAACGTGGTCAAGCAGGATGGCCGCGACGTTTCATATCGTCATCAAGCCCGATCATGAATG  
CGTGGTCAATTCCGGTCTGGGTTTCAGTGC GCGGCGCCAGAATGGCCGAAACAAGTTTTTCCGAAGCGCGC  
AACACCCAGCAACAGCGTCTGACAGATCCGCTCGTCTGGCGCTACGGCCAGATGCAGCCGACGAGCTGGG  
ACGACGCACTCGATCTTGTGCGCCGCGTGACAGCAAGGATTGTTAAAGACAAGGGTGAGGACGCACTGAT  
TGTTTCTGCCTTCGATCATGGCGGAGCTGGCGGCGGCTATGAAAACACCTGGGGCACGGGCAAGCTTTAT  
TTCGAGGCCATGAAGGTCAAGAATATTTCGAATCCACAACCGCCCTGCCTACAATTCCGAGGTCCACGGCA  
CGCGCGACATGGGCGTCGGCGAATTGAACAATTGCTACGAGGACGCCGAACCTCGCCGATACGATCGTCGC  
GGTCGGAACCAATGCGCTCGAGACCCAAACCAATTATTTCTTAAACCACTGGGTCCCAAACCTGCGCGGC  
GAAAGCCTCAACAAGAAAAAGCAGGTACTACCGAACGAACCGCATGAGGCTGCCAGAATCATCATCGTCG  
ATCCGCGCCGCACTGTGACGATCAACGCCTGTGAGGAAGTGGCCGGCAAGGACAACGTCCTGCATTTAGC  
GATCAACTCCGGCACTGACCTCGCGCTGTTCAATGCACTGTTACATATATTGCGGACAAGGGCTGGGTT  
GATCGCGAATTCATCGACAAATCGACCTCGCGCCAAGGCGTGGCAAGGCCGCGCAATATCCCGCGCGTG  
GCGAATCGGACGCCAAACCGGGACATTTATCCAATTTCGAGGGCGCCGTGCAAGGCTGTCTGTATATCCAT  
TGAGGACGCCGCGAAAATAACGGGGCTCGAGCAAGCCCAGATCGTCAAAGCCGCGAGTGATCGGCATG  
CCGAAGGAAGGCGGAAAGCGCCGCGCGTCATGTTTGGTTACGAGAAGGGCCTCATCTGGGGCAACGACA

>KT992346.1 *Rhizobium* sp. strain K3 (aioA) gene, partial cds

TGTATTGTGGGGTGTGGTTATCACGCCTACACGTGGCCGATCAACAAGCAGGGCGGTACCGATCCTCAGA  
AGAACGTGTTTGGCGTCGATCTGTTCGGAACAACAGCAGGCGGAGACTGACGCCTGGTATTCGCCATCCAT  
GTACAACGTGGTCAAGCAGGATGGCCGCGACGTTTCATATTGTCATCAAGCCCGATCATGAATGCGTGGTC  
AATTCCGGCCTGGGCTCGGTACGCGGCGCCAGAATGGCCGAAACAAGTTTTTCCGAAGCGCGCAACACCC  
AGCAACAGCGTCTGACAGATCCGTTGGTCTGGCGCTACGGCCAGATGCAGCCGACGAGCTGGGACGATGC  
ACTCGATCTTGTGCGCCGCGTGACCGCAAAGATCGTGAAAGATAAGGGTGAGGACGCACTGATCGTTTCT  
GCCTTCGATCACGGAGGTGCTGGCGGCGGCTATGAAAACACCTGGGGCACCGGCAAGCTTTATTTTCGAGG  
CCATGAAGGTCAAGAATATTTCGAATGCACAACCGCCCTGCCTACAATTCCGAGGTCCACGGCACGCGCGA  
CATGGGCGTCGGTGAACGAACAATTGCTACGAGGACGCCGAACCTTGCCGATACGATCGTCGCGGTTCGGA  
ACCAACGCGCTCGAGACCCAAACCAACTATTTCTTGAACCACTGGGTCCCAAACCTTGCGTGGCGAAAGCC  
TCAACAAGAAAAAACAGGTACTGCCGAACGAACCGCATGAGGCAGCCAGAATCATCATCGTCGATCCGCG

CCGCACTGTGACGATCAACGCCTGTGAGGAAGTAGCCGGCAAGGACAACGTCCTGCATTTAGCGATCAAC  
TCCGGCACTGACCTCGCTCTGTTCAATGCACTGTTACATACATCGCGGACAAGGGCTGGGTTGATCGCG  
AATTCATCGACAAATCGACCTCTCGTCAAGGCGTGGCACGGCCGCCGAATATCCCGCACGTGGCGAATC  
CGACAGTAAGCCTGGACATTTGTCCAACCTCGAGGGCGCCGTGCAAGGCTGTCGTATATCCATTGAGGAC  
GCCGCGAAAATAACGGGGCTAGAGCAAGCCGAGATCGTCAAGGCAGCCGAGTGGATCGGCATGCCGAAGG  
AAGGCGGAAAGCGCCGCCGCGTGTGTTTCGGTTACGAAAAAGGCCTGATCTGGGGCAACGATAACTACCG  
CACCAACGGCGCCGTGGTGAACCTTGCCTGGCCACGGGCAATATTGGCCGACCGGGTGGCGGCGTCGTT  
CGTTTGGGCGGCCACCAAGAAGGATA

>KT992341.1 *Rhizobium* sp. strain 3AM-13 (aioA) gene, partial cds

CCACTTCTGCATCGTGGGATGTGGGTACCACGCCTACACGTGGCCGATCAACAAGCAGGGCGGTACCGAT  
CCTCAGAAGAATGTATTTGGCGTCGATCTGTGGAACAACAGCAGGCGGAGACTGACGCCTGGTATTTCG  
CATCCATGTACAACGTGGTCAAGCAGGATGGCCGCGACGTTTCATATCGTCATCAAGCCCGATCATGAATG  
CGTGGTCAATTCCGGCCTCGGCTCGGTGCGCGGCCAGAATGGCAGAAACGAGTTTTTCCGAAGCCCGC  
AACACCCAGCAACAGCGTCTGACAGATCCGTTGGTCTGGCGCTACGGCCAGATGCAGCCGACGAGCTGGG  
ACGATGCACTCGATCTTGTGCCCCGCGTGACCGCAAAGATCGTGAAAGATAAGGGTGAGGACGCACTGAT  
CGTTTCTGCCTTCGATCACGGAGGTGCTGGCGGCGGCTATGAAAACACCTGGGGCACGGGCAAGCTTTAT  
TTCGAGGCCATGAAGGTCAAGAATATTGCAATCCACAACCGCCCTGCCTACAATTCCGAGGTCCACGGCA  
CGCGCGACATGGGCGTCGGTGAATTGAACAATTGCTACGAGGACTCCGAACTTGCCGATACGATCGTCGC  
GGTCGGAACCAACGCGCTCGAGACCCAAACCAACTATTTCTGAACCACTGGGTCCCAAACCTGCGTGGC  
GAAAGCCTCAACAAGAAAAAACAGGTACTGCCGAACGAACCGCATGAGGCAGCCAGAATCATCATCGTCG  
ATCCGCGCCGCACTGTGACGATCAACGCCTGTGAGGAAGTAGCCGGCAAGGACAACGTCCTGCATTTAGC  
GATCAACTCCGGCACTGACCTCGCTCTGTTCAATGCACTGTTACATACATCGCGGACAAGAGCTGGGTT  
GATCGCGAATTCATCGACAAATCGACCTCTCGTCAGGGCGTGSCACGGCCGCCGAATATCCCGCACGTG  
GCGAATCCGACAGTAAACCTGGACATTTGTCCAACCTCGAGGGCGCCGTCAAGGGCTGTCGTATATCCAT  
TGAGGACGCCGCGAAAATAACGGGGCTAGAGCAAGCCGAGATCGTCAAAGCCGCCGAGTGGATCGGCATG  
CCGAAGAAAGGCGGAAAGCGCCGCCGCGTGTGTTTCGGTTACGAGAAAGGCCTCATCTGGGGCAACGACA

>MF621580.1 *Rhizobium* sp. strain Cug7 (aioA) gene, partial cds

ATTCCAATTCTGCATCGTGGGTTGTGGGTATCACGCCTACACGTGGCCGATCAACAAGCAGGGCGGTACC  
GATCCTCAGAAGAACGTGTTTGGCGTCGATCTGTGGAACAACAGCAGGCGGAGACTGACGCCTGGTATT

CGCCATCCATGTACAACGTGGTCAAGCAGGATGGCCGCGACGTTTCATATTGTCATCAAGCCCGATCATGA  
ATGCGTGGTCAATTCCGGCCTGGGCTCGGTACGCGGCGCCAGAATGGCCGAAACAAGTTTTTCCGAAGCG  
CGGAACACCCAGCAACAGCGTCTGACAGATCCGTTGGTCTGGCGCTACGGGCAGATGCAGCCGACGAGCT  
GGGACGACGCACTCGATCTTGTGCGCCGCGTGACCGCAAGGATCGTGAAAGATAAGGGTGAGGACGCCCT  
GATCGTTTTCTGCCTTTGATCATGGCGGTGCTGGCGGCGGCTATGAGAACACCTGGGGCACCGGCAAGCTT  
TATTTTCGAGGCCATGAAGGTCAAGAATATTCGAATCCACAACCGCCCTGCCTACAATTCCGAGGTGCACG  
GCACGCGCGACATGGGCGTCGGTGAGTTGAACAACCTGCTACGAAGATGCCGAACCTTGCCGATACGATCGT  
CGCGATCGGTACCAACGCGCTTGAGACCCAAACCAATTATTTCTGAACCACTGGGTCCCGAACCTTGCGC  
GGCGAAAGCCTAAACAAGAAAAAGCAGATACTGCCGAACGAACCGCATGAGGCAGCCAGGATCATCATCG  
TTGATCCGCGCCGCACTGTGACGATCAACGCCTGTGAGGAAGTAGCCGGCAAGGACAACGTCCTGCATTT  
AGCGATCAACTCCGGCACTGACCTCGCTCTGTTCAATGCACTGTTTACATACATCGCGGACAAGGGCTGG  
GTTGATCGCGAATTCATCGACAAATCGACCTTGCGCCAAAGCTTGGCAAGGCCGCCGAATATCCCGCGC  
GTGGCGAATCGGACGCCAAACCGGGACATTTATCCAACCTTCGAGGGCGCCGTCGAAGGCTGTCTATATC  
CATTGAGGACGCCGCGAAAATAACGGGGCTCGAGCAAGCCAGATCGTCAAAGCCGCCGAGTGGATCGGC  
ATGCCGAAGGAAGGCGGAAAGCGCCGCCGCGTCATGTTTGGTTACGAAAAGGAATCATCTGGGGCAACG  
ACAAATCGTCGAC

>MF621579.1 *Rhizobium* sp. strain Cug6 (aioA) gene, partial cds

ATTCCACTTCTGCATCGTGGGGTGTGGTTACAAGGCCTACACCTGGGACATCAACCGGCAGGGCGGCACG  
AATCCGAGCCAGAACAAGTTCAAGGCCGATCTTTCCAAGCAGCAGGGTGCCGACAGCAGCGCCTGGTACA  
CGCCGTCGATGTACAACATCGTCAAGCAGGACGGTAAGGACGTCCACCTGGTCATCATGCCGACAAGAA  
CTGCGTGGTGAATTCGGGCCTCGGTTCCGTCGCGGCGCCCGCATGGCGGAGACCTCGTACTCAGAGGCG  
CGTTCGACCCAACAGCAGCGCCTCACGCACCCGATGGTCTGGCGCTATGGAGCGATGTCTCCGACGTCCT  
GGGATGATGCTCTCGATCTCGTGGCCCGGGTCACCTGCCAGATCGTCAAGGATCAGGGCGAAGATGGGCT  
TTTCGTCTCCGCTTTTCGATCATGGCGGCGCCGGCGGCGGCTACGAGAACAACCTGGGGCACCGGCAAGCTC  
TATTTTCGGCGCCATGAAGGTTAAGAACATCCGGATCCACAACCGCCGGCCTACAATTCGGAGGTCCACG  
CCACGCGCGACATGGGCATCGGCGAGCTCAACAACCTGCTACGAGGATGCCGAGCTTGCGGACACGATCGT  
CGTGGTAGGCGCCAATCCGCTGGAGACCCAGACCAACTATTTCTGAACCATTGGGTGCCGAACCTTGCGC  
GGCACGTGACGGACAAGAAGCGCGCCGAGCTTCCGAATGAGGCTCATCCACCGGCACGGATCGTCATCA  
TCGATCCGCGGCGCACAGTCACGGTCAACGCCTGCGAGGTGGAAGCCGGCAAGGACCGGGTGATGCACCT  
CGCTATCAACTCGGGCAGCGATCTTGCGCTGTTCAATGCCTGGATGACGTACATTGCCGAGAAGGGTTGG

GTCGACAGGGCGCTGATCGCGGCTCGACCAACGGGTTGACAAGATGGTCGCGGCCAACAAGACGACCC  
TGGAGCAGGCTGCCGCTCTGACGGGCCTCACCGTCGACCAGATCCGCCAGTCGGCGGAATGGATCGCGTC  
ACCAAAGGAAGGCAATGCGCGACGCCGCACCATGTTTCGCCTATGAAAAAGGTATCATCTGGGGCAACGAC  
AAATCGTCGAC

>KT992342.1 *Shinella* sp. strain C23 (aioA) gene, partial cds

CCACTTCTGCATCGTGGGGTGCGGGTACCACGCCTATACGTGGCCGATCAACAAGCAGGGCGGTACCGAT  
CCGCAGAAGAACGTGTTTGGCGTCGATCTGTCGGAACAACAGCAGGCGGAAACCGATGCCTGGTATTTCG  
CATCCATGTACAACGTGGTCAAGCAGGATGGCCGCGATGTTTCATATTGTCATCAAGCCCGACCATGAATG  
CGTGGTCAATTCCGGCCTCGGCTCGGTGCGCGGCGCAAGAATGGCCGAAACGAGTTTTTCCGAAGCGCGC  
AACACCCAGCAACAGCGTCTGACAGATCCGTTGGTCTGGCGCTACGGCCAGATGCAGCCGACAAGCTGGG  
ACGATGCACTCGATCTTGTGCGACGCGTGACCTCGAGGATCGTGAAAGACAAGGGTGAGGACGCGCTGAT  
TGTTTCTGCCTTCGATCATGGAGGTGCTGGCGGCGGCTATGAAAACACCTGGGGCACCGGCAAGCTTTAT  
TTCGAAGCGATGAAGGTCAAGAATATTTCGCATCCACAACCGCCCTGCCTACAATTCCGAGGTGCACGGA  
CGCGCGACATGGGCGTCGGTGAATTGAACAACTGCTACGAAGACGCCGAACCTTGCCGATACGATCGTGGC  
GATCGGGACCAACGCGCTCGAGACGCAAACCAACTATTTCTGAACCATTGGGTCCCAAACCTTGCGGGGC  
GAAAGCCTTAACAAAAAAGCAGGTGATGCCCAACGAACCGCATGAGGCAGCCAGGATCATTATAGTCG  
ATCCGCGCCGCACAGTGACGATCAACGCCTGCGAAGAAGTAGCTGGCAAGGATAATGTTCTGCATCTTGC  
GATCAACTCCGGCACTGACCTCGCTCTGTTCAATGCACTGTTACCTATATCGCCGACAAGGGCTGGGTC  
GATCGGGAATTCATCGACAAATCGACCTTTCGTCAACGCGTGGCACGGCCGCCGCAATATCCCGCGAGCG  
GCGAATCGGACAGCCAGCCTGGACATTTGTCCAACCTTCGAGGGCGCCGTCGAAGGCTGTCGTATGTCCAT  
CGAGGATGCCCGGAAAATAACGGGGCTCGAGCAAGCCGAGATCGTCAAGGCAGCCGAGTGGATCGGTATG  
CCGAAGGAAGGCGGAAAGCGCCGCCGAGTGATGTTTCGGTTACGAAAAGGGCCTCATCTGGGGCAACGACA

>MF621576.1 *Shinella* sp. strain Cug3 (aioA) gene, partial cds

ATTCCACTTCTGCATCGTGGGGTGCGGGTATCACGCATATACCTGGCCGGCAAACAAGCAGGGTGGTACA  
GCGCCCCGACCGCAATGTTTTCGGCGTTAACCTGTCAGAACAGCAATATGCGGAATCGGACGCCTGGTATT  
CGCCCTCCATGTACAACATGGTGAAGCAGGACGGTCAGGACGTTACATTGTTATCAAGCCGGACCAGGC  
TTGCGTCGTCAATTCGGGGCTCGGCTCCGTGCGCGGCGCGGTATGGCGGAGATGAGTTTCTCCGAGGCG  
CGCAACACACAGCAGCAGCGCCTGACCGACCCGCTTGTCTGGCGTTACGGCCAGATGCAGCCGACAAGCT  
GGGAAGATGCCCTTGACCTTGTGGCGCGCGTTACCGCGGCGGTTCGTCAAAGCGAAGGGCGAGGATGCCCT

CTATGTCTCCGCCTATGATCACGGCGGCGCCGGCGGCGGCTATGAGAATACCTGGGGCACGGGCAAGCTC  
TACTTCGAGGCGATGAAGATCAAGAATATCCGCATCCACAATCGCCCGGCATACAACCTCGGAAGTCCATG  
CAACGCGGGACATGGGGGTGGGCGAGCTCAACAATTGCTACGAGGACGCAGAACTTGCTGACACGATCTT  
CTGCGTTGGTGCAAATCCGCTGGAAACCCAGACAACTACTTCCTCAACCACTGGATTCCCAATCTTCGC  
GGGACGACCGCAGCGAAAAAAGACGAGCAGTTGCCGGGTGAAGAGCACCCGCCTGCCCCGAATTGTCTTTG  
TAGACCCGCGTCGTACCGTCAGCATCAACGCGGCGGAGGCAGAAGCCGGTAAAGACAACGTCCTGCATCT  
TGCGATCAATTCAAGCACCGACCTTGCTCTGTTCAACGCCATTTTCAGCCATATCGTCGAAAAGGGCTGG  
GTGGATCAGTCGTTTATCGATGCCAGCACCGCCGGCTTCACGGAGGCTGTTGCAGCCAACAAGTTGACGC  
TCGAGAAGGCCAGTGAGATTTGCGGCGTCTCGGCTTCCGATATCGCGAAGGCGGCACAGTGGATTGCCGA  
ACCCAAGGCCGACGGCAAGCGCCGACGCACCATGCTTCCCTATGAAAAAGGTTTCATCTGGGGCAACGAC  
AAATCGTCGAC

>AB915361.1 Sinorhizobium sp. KGO-5 (aioA) gene, partial cds

AGGGAGGAACCGAACCCGACCGCAACGCCTTCGCCGTCCCACCTCTCGAAGCAGCAAGAGGCGGAGACCG  
ACGCCTGGTACTCGCCATCGATGTACAACATCGTCAAGCAAGATGGGCAGGACGTTTCATATCGTCATCAA  
GCCGGACCACGAGTGCGTGGTCAATTCGGGCCTTGGCTCCGTCCGCGGCGCGCGTATGGCCGAGATGAGC  
TACTCCGAGGCCCCGAATACACAGCAGCAGCGCCTGACCGATCCCCTCGTATGGAGATACGGACAGATGC  
AGCCGACGAGTTGGGACGACGCATTGGATCTAGTGGCACGCGTGACGGCGACAATAGTCAAGGAAAAGGG  
CGAGGATGCCCTCTACGTCTCGGCCTACGACCATGGCGGCGCGGGCGGTGGCTACGAGAACACCTGGGGT  
ACGGGCAAGCTCTATTTTCGAGGCGATGAAAGTCAAGAATATCCGTATCCACAACCGCCCTGCCTACAAC  
CCGAAGTCCACGCCACCCGGGACATGGGCGTGGCGAACTTAACAACCTGTTACGAGGATGCCGAAC  
TAGCAGACGATCTTCTGCGTCGGTGCTAATCCCCTTGAGACACAGACGAACTATTTCTCAACCACTGGGTC  
CCAAACCTGCGGGGATCGACTGCGGACAAGAAGAAGCAGCAGCTTCCAGGTGAGGAGCATCCGCCAGCAC  
GCATCGTGTTTCGTCGATCCACGACGCACAGTGACCATCAACGCAGCCGAAGCCGATGCGGGCCTGGAAAA  
CGTCCTGCATCTCGCGATCAACTCTGGCACTGACCTTGCTCTGTTCAACGCGATCTTCACGCACATCGTC  
GAGAAGGGCTGGATCGACAAGGCGTTCATCGACGCCAGCACAAAGCGGA

b) So sánh nhiều trình tự bằng chương trình Clustal Omega của EBI

Copy kết quả so sánh nhiều trình tự vào ô trống dưới đây

CLUSTAL O(1.2.4) multiple sequence alignment

KX250215.1	-----	0
MF621579.1	ATTCCACTTCTGCATCGTGGGGTGTGGTTACAAGGCCTACACCTGGGACATCAACCGGCA	60
KT992342.1	---CCACTTCTGCATCGTGGGGTGCGGGTACCACGCCTATACGTGGCCGATCAACAAGCA	57
KT992346.1	-----TGTATTGTGGGGTGTGGTTATCACGCCTACACGTGGCCGATCAACAAGCA	50
KT992341.1	---CCACTTCTGCATCGTGGGATGTGGGTACCACGCCTACACGTGGCCGATCAACAAGCA	57
KT992350.1	---CCACTTCTGCATCGTGGGGTGTGGGTACCACGCCTACACGTGGCCGATCAACAAGCA	57
MF621580.1	ATTCCACTTCTGCATCGTGGGGTGTGGGTATCACGCCTACACGTGGCCGATCAACAAGCA	60
MF621576.1	ATTCCACTTCTGCATCGTGGGGTGCGGGTATCACGCATATACCTGGCCGGCAAACAAGCA	60
AB915361.1	-----A	1

KX250215.1	-----GACAACGCGTTCGGCGTCGACCTCGCCGAGCAGCAGGGGG	40
MF621579.1	GGGCGGCACGAATCCGAGCCAGAACAAG-TTCAAGGCCGATCTTTCCAAGCAGCAGGGTG	119
KT992342.1	GGGCGGTACCGATCCGCAGAAGAACGTG-TTTGGCGTCGATCTGTCGGAACAACAGCAGG	116
KT992346.1	GGGCGGTACCGATCCTCAGAAGAACGTG-TTTGGCGTCGATCTGTCGGAACAACAGCAGG	109
KT992341.1	GGGCGGTACCGATCCTCAGAAGAATGTA-TTTGGCGTCGATCTGTCGGAACAACAGCAGG	116
KT992350.1	GGGCGGTACCGATCCTCAGAAGAACGTG-TTTGGCGTCGATCTGTCGGAACAACAGCAGG	116
MF621580.1	GGGCGGTACCGATCCTCAGAAGAACGTG-TTTGGCGTCGATCTGTCGGAACAACAGCAGG	119
MF621576.1	GGGTGGTACAGCGCCCGACCGCAATGTT-TTCGGCGTTAACCTGTCAGAACAGCAATATG	119
AB915361.1	GGGAGGAACCGAACCCGACCGCAACGCCTTCGCCGTCCACCTCTCGAAGCAGCAAGAGG	61

\* \* \* \* \*

KX250215.1	CGGACACCCCGCCTGGTACGCCCCGTCCATGTACAACATTGTCCGCCAGGACGGCCGAG	100
MF621579.1	CCGACAGCAGCGCCTGGTACACGCCGTCGATGTACAACATCGTCAAGCAGGACGGTAAGG	179
KT992342.1	CGGAAACCGATGCCTGGTATTCGCCATCCATGTACAACGTGGTCAAGCAGGATGGCCGCG	176
KT992346.1	CGGAGACTGACGCCTGGTATTCGCCATCCATGTACAACGTGGTCAAGCAGGATGGCCGCG	169
KT992341.1	CGGAGACTGACGCCTGGTATTCGCCATCCATGTACAACGTGGTCAAGCAGGATGGCCGCG	176
KT992350.1	CGGAGACTGACGCCTGGTATTCGCCATCCATGTACAACGTGGTCAAGCAGGATGGCCGCG	176
MF621580.1	CGGAGACTGACGCCTGGTATTCGCCATCCATGTACAACGTGGTCAAGCAGGATGGCCGCG	179
MF621576.1	CGGAATCGGACGCCTGGTATTCGCCCTCCATGTACAACATGGTGAAGCAGGACGGTCAGG	179
AB915361.1	CGGAGACCGACGCCTGGTACTCGCCATCGATGTACAACATCGTCAAGCAAGATGGGCAGG	121

\* \* \* \* \*

KX250215.1	ATGTGACATCGTCATCAAGCCCGACCGCGAATGCTCGGTGAACTCCGGCCTCGGATCGG	160
MF621579.1	ACGTCCACCTGGTCATCATGCCGACAAGAACTGCGTGGTGAATTCCGGCCTCGGTTCGG	239
KT992342.1	ATGTTTCATATTGTCATCAAGCCCGACCATGAATGCGTGGTCAATTCCGGCCTCGGCTCGG	236
KT992346.1	ACGTTTCATATTGTCATCAAGCCCGATCATGAATGCGTGGTCAATTCCGGCCTGGGCTCGG	229
KT992341.1	ACGTTTCATATCGTCATCAAGCCCGATCATGAATGCGTGGTCAATTCCGGCCTCGGCTCGG	236
KT992350.1	ACGTTTCATATCGTCATCAAGCCCGATCATGAATGCGTGGTCAATTCCGGTCTGGGTTTCAG	236
MF621580.1	ACGTTTCATATTGTCATCAAGCCCGATCATGAATGCGTGGTCAATTCCGGCCTGGGCTCGG	239
MF621576.1	ACGTTTCACATTGTTATCAAGCCCGACCGAGCTTGCCTCGTCAATTCCGGCCTCGGCTCCG	239
AB915361.1	ACGTTTCATATCGTCATCAAGCCCGACCGAGTGCCTGGTCAATTCCGGCCTTGGCTCCG	181

\* \*

KX250215.1	TGCGCGGCGCGCGCATCGCGGAGATGAGCTATTCGCGCCAACGCAACACGCAGCTCCAGC	220
MF621579.1	TCCGCGGCGCCCGCATGGCGGAGACCTCGTACTCAGAGGCGGTTTCGACCCAACAGCAGC	299
KT992342.1	TGCGCGGCGCAAGAATGGCCGAAACGAGTTTTTCCGAAGCGCGCAACACCCAGCAACAGC	296
KT992346.1	TACGCGGCGCCAGAATGGCCGAAACAAGTTTTTCCGAAGCGCGCAACACCCAGCAACAGC	289
KT992341.1	TGCGCGGCGCCAGAATGGCAGAAACGAGTTTTTCCGAAGCCCGCAACACCCAGCAACAGC	296
KT992350.1	TGCGCGGCGCCAGAATGGCCGAAACAAGTTTTTCCGAAGCGCGCAACACCCAGCAACAGC	296
MF621580.1	TACGCGGCGCCAGAATGGCCGAAACAAGTTTTTCCGAAGCGCGCAACACCCAGCAACAGC	299
MF621576.1	TGCGCGGCGCGCGTATGGCGGAGATGAGTTTCTCCGAGGCGCGCAACACACAGCAGCAGC	299
AB915361.1	TCCGCGGCGCGCGTATGGCCGAGATGAGTACTCCGAGGCCCGCAATACACAGCAGCAGC	241

\* \*

KX250215.1	GCCTTACCGATCCGCTGGTCTGGCGCTACGGCCAGCTGCAACCGACCAGCTGGGACGACG	280
MF621579.1	GCCTCACGCACCCGATGGTCTGGCGCTATGGAGCGATGTCTCCGACGTCCTGGGATGATG	359
KT992342.1	GTCTGACAGATCCGTTGGTCTGGCGCTACGGCCAGATGCAGCCGACAAGCTGGGACGATG	356
KT992346.1	GTCTGACAGATCCGTTGGTCTGGCGCTACGGCCAGATGCAGCCGACGAGCTGGGACGATG	349
KT992341.1	GTCTGACAGATCCGTTGGTCTGGCGCTACGGCCAGATGCAGCCGACGAGCTGGGACGATG	356
KT992350.1	GTCTGACAGATCCGCTCGTCTGGCGCTACGGCCAGATGCAGCCGACGAGCTGGGACGACG	356
MF621580.1	GTCTGACAGATCCGTTGGTCTGGCGCTACGGGCGAGATGCAGCCGACGAGCTGGGACGACG	359
MF621576.1	GCCTGACCGACCCGCTTGTCTGGCGTTACGGCCAGATGCAGCCGACAAGCTGGGAAGATG	359
AB915361.1	GCCTGACCGATCCCTCGTATGGAGATACGGACAGATGCAGCCGACGAGTTGGGACGACG	301

\* \*

KX250215.1	CGCTCGACCTCGTCGCCCCGGGTGACATCGGCCGTGATCGCCGAACAGGGCGAGGATGGGC	340
MF621579.1	CTCTCGATCTCGTGGCCCCGGGTACCTGCCAGATCGTCAAGGATCAGGGCGAAGATGGGC	419
KT992342.1	CACTCGATCTTGTGCGACGCGTGACCTCGAGGATCGTGAAAGACAAGGGTGAGGACGCGC	416
KT992346.1	CACTCGATCTTGTGCGCCCGCGTGACCGCAAAGATCGTGAAAGATAAGGGTGAGGACGCAC	409
KT992341.1	CACTCGATCTTGTGCGCCCGCGTGACCGCAAAGATCGTGAAAGATAAGGGTGAGGACGCAC	416
KT992350.1	CACTCGATCTTGTGCGCCCGCGTGACAGCAAGGATTGTTAAAGACAAGGGTGAGGACGCAC	416
MF621580.1	CACTCGATCTTGTGCGCCCGCGTGACCGCAAGGATCGTGAAAGATAAGGGTGAGGACGCCC	419
MF621576.1	CCCTTGACCTTGTGGCGCGCGTTACCGCGCGGTCGTCAAAGCGAAGGGCGAGGATGCCC	419
AB915361.1	CATTGGATCTAGTGGCACGCGTGACGGCGACAATAGTCAAGGAAAAGGGCGAGGATGCCC	361

\* \* \* \* \*

KX250215.1	TGTTTCGTCCTCCGCTTTTCGACCA	400
MF621579.1	TTTTCGTCTCCGCTTTTCGATCATGGCGGCGCCGGCGGGCTACGAGAACACTTGGGGCA	479
KT992342.1	TGATTGTTTCTGCCTTCGATCATGGAGGTGCTGGCGGCGGCTATGAAAACACCTGGGGCA	476
KT992346.1	TGATCGTTTCTGCCTTCGATCACGGAGGTGCTGGCGGCGGCTATGAAAACACCTGGGGCA	469
KT992341.1	TGATCGTTTCTGCCTTCGATCACGGAGGTGCTGGCGGCGGCTATGAAAACACCTGGGGCA	476
KT992350.1	TGATTGTTTCTGCCTTCGATCATGGCGGAGCTGGCGGCGGCTATGAAAACACCTGGGGCA	476
MF621580.1	TGATCGTTTCTGCCTTTGATCATGGCGGTGCTGGCGGCGGCTATGAGAACACCTGGGGCA	479
MF621576.1	TCTATGTCTCCGCCTATGATCACGGCGGCGCCGGCGGCGGCTATGAGAATACCTGGGGCA	479
AB915361.1	TCTACGTCTCGGCCTACGACCATGGCGGCGCGGGCGGTGGCTACGAGAACACCTGGGGTA	421

\* \* \* \* \*

KX250215.1	CCGGCAAGCTCTATTTTGGGGCGATGAAGGTCAAGAACATCCGCATCCACAACCGCCCGG	460
MF621579.1	CCGGCAAGCTCTATTTTCGGCGCCATGAAGGTAAAGAACATCCGGATCCACAACCGGCCGG	539
KT992342.1	CCGGCAAGCTTTATTTTCGAAGCGATGAAGGTCAAGAATATTCGCATCCACAACCGCCCTG	536
KT992346.1	CCGGCAAGCTTTATTTTCGAGGCCATGAAGGTCAAGAATATTCGAATGCACAACCGCCCTG	529
KT992341.1	CGGGCAAGCTTTATTTTCGAGGCCATGAAGGTCAAGAATATTCGAATCCACAACCGCCCTG	536
KT992350.1	CGGGCAAGCTTTATTTTCGAGGCCATGAAGGTCAAGAATATTCGAATCCACAACCGCCCTG	536
MF621580.1	CCGGCAAGCTTTATTTTCGAGGCCATGAAGGTCAAGAATATTCGAATCCACAACCGCCCTG	539
MF621576.1	CGGGCAAGCTCTACTTCGAGGCGATGAAGATCAAGAATATCCGCATCCACAATCGCCCGG	539
AB915361.1	CGGGCAAGCTCTATTTTCGAGGCGATGAAAGTCAAGAATATCCGTATCCACAACCGCCCTG	481

\* \* \* \* \*

KX250215.1	CCTACAATTCCGAGGTTCACGGCTCGCGCGACATGGGCGTCGGCGAACTGAACAATTGCT	520
MF621579.1	CCTACAATTCCGAGGTCCACGCCACGCGCGACATGGGCATCGGCGAGCTCAACAACCTGCT	599
KT992342.1	CCTACAATTCCGAGGTGCACGGAACGCGCGACATGGGCGTCGGTGAATTGAACAACCTGCT	596
KT992346.1	CCTACAATTCCGAGGTCCACGGCACGCGCGACATGGGCGTCGGTGAACCTGAACAATTGCT	589
KT992341.1	CCTACAATTCCGAGGTCCACGGCACGCGCGACATGGGCGTCGGTGAATTGAACAATTGCT	596
KT992350.1	CCTACAATTCCGAGGTCCACGGCACGCGCGACATGGGCGTCGGCGAATTGAACAATTGCT	596
MF621580.1	CCTACAATTCCGAGGTGCACGGCACGCGCGACATGGGCGTCGGTGAGTTGAACAACCTGCT	599
MF621576.1	CATACAACCTCGGAAGTCCATGCAACGCGGGACATGGGGGTGGGCGAGCTCAACAATTGCT	599
AB915361.1	CCTACAACCTCCGAAGTCCACGCCACCCGGGACATGGGCGTCGGCGAACTTAACAACCTGTT	541

\* \* \* \* \*

KX250215.1	ATGAGGACGCCAGCTCGCCGACACCATCGTGGCCGTCGGCACCAACGCGCTGGAAACCC	580
MF621579.1	ACGAGGATGCCGAGCTTGCGGACACGATCGTCGTGGTAGGCGCCAATCCGCTGGAGACCC	659
KT992342.1	ACGAAGACGCCGAACCTTGCCGATACGATCGTGGCGATCGGGACCAACGCGCTCGAGACGC	656
KT992346.1	ACGAGGACGCCGAACCTTGCCGATACGATCGTCGCGGTTCGGAACCAACGCGCTCGAGACCC	649
KT992341.1	ACGAGGACTCCGAACCTTGCCGATACGATCGTCGCGGTTCGGAACCAACGCGCTCGAGACCC	656
KT992350.1	ACGAGGACGCCGAACCTCGCCGATACGATCGTCGCGGTTCGGAACCAATGCGCTCGAGACCC	656
MF621580.1	ACGAAGATGCCGAACCTTGCCGATACGATCGTCGCGATCGGTACCAACGCGCTTGAGACCC	659
MF621576.1	ACGAGGACGCAGAACTTGCTGACACGATCTTCTGCGTTGGTGCAAATCCGCTGGAAACCC	659
AB915361.1	ACGAGGATGCCGAACCTAGCGGACACGATCTTCTGCGTCGGTGCTAATCCCTTGAGACAC	601

\* \* \* \* \*

KX250215.1	AGACCAACTACTTCCTGAACCATTGGGTGCCGAATTTGCGCGGCACCTCGATGGACAAAA	640
MF621579.1	AGACCAACTATTTCTGAACCATTGGGTGCCGAACCTGCGCGGCACGTCGACGGACAAGA	719
KT992342.1	AAACCAACTATTTCTGAACCATTGGGTCCCAAACCTGCGGGGCGAAAGCCTTAACAAAA	716
KT992346.1	AAACCAACTATTTCTGAACCACTGGGTCCCAAACCTTGCCTGGCGAAAGCCTCAACAAGA	709
KT992341.1	AAACCAACTATTTCTGAACCACTGGGTCCCAAACCTTGCCTGGCGAAAGCCTCAACAAGA	716
KT992350.1	AAACCAATTATTTCTAAACCACTGGGTCCCAAACCTGCGCGGCGAAAGCCTCAACAAGA	716
MF621580.1	AAACCAATTATTTCTGAACCACTGGGTCCCGAACTTGCCTGGCGGCGAAAGCCTAAACAAGA	719
MF621576.1	AGACAAACTACTTCCTCAACCACTGGATTCCCAATCTTCGCGGGACGACCGCAGCGAAAA	719
AB915361.1	AGACGAACCTATTTCTCAACCACTGGGTCCCAAACCTGCGGGGATCGACTGCGGACAAGA	661

\* \* \* \* \*

KX250215.1	AGAAGGCCGAGTTCGGCAGCGAGCCGGTGGCTCAGGGCCGGGTGGTCATCGTCGGTCCGC	700
MF621579.1	AGCGCGCCGAGCTTCCGAATGAGGCTCATCCACCGGCACGGATCGTCATCATCGATCCGC	779
KT992342.1	AAAAGCAGGTGATGCCCCAACGAACCGCATGAGGCAGCCAGGATCATTATAGTCGATCCGC	776
KT992346.1	AAAAACAGGTACTGCCGAACGAACCGCATGAGGCAGCCAGAATCATCATCGTCGATCCGC	769
KT992341.1	AAAAACAGGTACTGCCGAACGAACCGCATGAGGCAGCCAGAATCATCATCGTCGATCCGC	776
KT992350.1	AAAAGCAGGTACTACCGAACGAACCGCATGAGGCTGCCAGAATCATCATCGTCGATCCGC	776
MF621580.1	AAAAGCAGATACTGCCGAACGAACCGCATGAGGCAGCCAGGATCATCATCGTTGATCCGC	779
MF621576.1	AAGACGAGCAGTTGCCGGGTGAAGAGCACCCGCCTGCCCGAATTGTCTTTGTAGACCCGC	779
AB915361.1	AGAACGAGCAGCTTCCAGGTGAGGAGCATCCGCCAGCACGCATCGTGTTCGTCGATCCAC	721

\* \* \*\* \* \* \* \* \*

KX250215.1	GCCGCACAGTGACGGTCAATGCCTGCGAAGTGGAGGCGGCAAGGACAACGTGCTTCATC	760
MF621579.1	GGCGCACAGTCACGGTCAACGCCTGCGAGGTGGAAGCCGGCAAGGACCGGGTGATGCACC	839
KT992342.1	GCCGCACAGTGACGATCAACGCCTGCGAAGAAGTAGCTGGCAAGGATAATGTTCTGCATC	836
KT992346.1	GCCGCACTGTGACGATCAACGCCTGTGAGGAAGTAGCCGGCAAGGACAACGTCCTGCATT	829
KT992341.1	GCCGCACTGTGACGATCAACGCCTGTGAGGAAGTAGCCGGCAAGGACAACGTCCTGCATT	836
KT992350.1	GCCGCACTGTGACGATCAACGCCTGTGAGGAAGTGGCCGGCAAGGACAACGTCCTGCATT	836
MF621580.1	GCCGCACTGTGACGATCAACGCCTGTGAGGAAGTAGCCGGCAAGGACAACGTCCTGCATT	839
MF621576.1	GTCGTACCGTCAGCATCAACGCGGCGAGGCAGAAGCCGGTAAAGACAACGTCCTGCATC	839
AB915361.1	GACGCACAGTGACCATCAACGCAGCCGAAGCCGATGCGGGCCTGGAAAACGTCCTGCATC	781

\* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \*

KX250215.1	TCGCGCTCAATTCCGGCACCGACCTCGCGTTGTTCAACGCCTGGTTCACCTATGCGGCGG	820
MF621579.1	TCGCTATCAACTCGGGCAGCGATCTTGCGCTGTTCAATGCCTGGATGACGTACATTGCCG	899
KT992342.1	TTGCGATCAACTCCGGCACTGACCTCGCTCTGTTCAATGCACTGTTACCTATATCGCCG	896
KT992346.1	TAGCGATCAACTCCGGCACTGACCTCGCTCTGTTCAATGCACTGTTACATACATCGCGG	889
KT992341.1	TAGCGATCAACTCCGGCACTGACCTCGCTCTGTTCAATGCACTGTTACATACATCGCGG	896
KT992350.1	TAGCGATCAACTCCGGCACTGACCTCGCGCTGTTCAATGCACTGTTACATATATTGCGG	896
MF621580.1	TAGCGATCAACTCCGGCACTGACCTCGCTCTGTTCAATGCACTGTTACATACATCGCGG	899
MF621576.1	TTGCGATCAATTGAGGCACCGACCTTGCTCTGTTCAACGCCATTTTCAGCCATATCGTCG	899
AB915361.1	TCGCGATCAACTCTGGCACTGACCTTGCTCTGTTCAACGCGATCTTCACGCACATCGTCG	841

\* \* \* \* \* \* \* \* \* \* \* \* \* \* \*

KX250215.1	AAAAGGGCTGGATCGACCGGAATTCATCGGCGCGTCCACGAAAGACTTCGACAAGGCGG	880
MF621579.1	AGAAGGGTTGGGTCGACAGGGCGCTGATCGCGGCCTCGACCAACGGGTTGACAAGATGG	959
KT992342.1	ACAAGGGCTGGGTCGATCGGAATTCATCGACAAATCGACCTTGCGTCAACGCGTGGCAC	956
KT992346.1	ACAAGGGCTGGGTTGATCGCGAATTCATCGACAAATCGACCTCTCGTCAAGGCGTGGCAC	949
KT992341.1	ACAAGAGCTGGGTTGATCGCGAATTCATCGACAAATCGACCTCTCGTCAGGGCGTGSCAC	956
KT992350.1	ACAAGGGCTGGGTTGATCGCGAATTCATCGACAAATCGACCTCGCGCCAAGGCGTGGCAC	956
MF621580.1	ACAAGGGCTGGGTTGATCGCGAATTCATCGACAAATCGACCTTGCGCCAAAGCTTGGCAC	959
MF621576.1	AAAAGGGCTGGGTGGATCAGTCGTTTATCGATGCCAGCACCGCCGGCTTCACGGAGGCTG	959
AB915361.1	AGAAGGGCTGGATCGACAAGGCGTTTCATCGACGCCAGCACAAAGCGGA-----	888
	* * * * * * * * * * * * * * * *	
KX250215.1	TCGCCGCCAACAAGGTCTCGATTGCCGAGCGGCGGAGATCACCGGCCTTTCCGAG----	936
MF621579.1	TCGCGGCCAACAAGACGACCTGGAGCAGGCTG-----	992
KT992342.1	GGCCGCCGAATATCCCGCGAGCGGCGAATCGGACAGCCAGCCTGGACATTTGTCCAAC	1016
KT992346.1	GGCCGCCGAATATCCCGCACGTGGCGAATCCGACAGTAAGCCTGGACATTTGTCCAAC	1009
KT992341.1	GGCCGCCGAATATCCCGCACGTGGCGAATCCGACAGTAACCTGGACATTTGTCCAAC	1016
KT992350.1	GGCCGCCGAATATCCCGCGCGTGGCGAATCGGACGCCAAACCGGGACATTTATCCAAC	1016
MF621580.1	GGCCGCCGAATATCCCGCGCGTGGCGAATCGGACGCCAAACCGGGACATTTATCCAAC	1019
MF621576.1	TTGCAGCCAACAAGTTGACGCTCG-AG--AAGGCCAGTGAGATTTGCG-----	1004
AB915361.1	-----	888
KX250215.1	-----GCCGACATCGTCAAGGCGATTACCTGGATCGCCGAGCCCA-----	977
MF621579.1	-----CCGCTCTGACGGGCC	1007
KT992342.1	TCGAGGGCGCCGTCTGAAGGCTGTCTGATGTCCATCGAGGATGCCGCGAAAATAACGGGGC	1076
KT992346.1	TCGAGGGCGCCGTCTGAAGGCTGTCTGATGTCCATCGAGGATGCCGCGAAAATAACGGGGC	1069
KT992341.1	TCGAGGGCGCCGTCTGAAGGCTGTCTGATGTCCATCGAGGATGCCGCGAAAATAACGGGGC	1076
KT992350.1	TCGAGGGCGCCGTCTGAAGGCTGTCTGATGTCCATCGAGGATGCCGCGAAAATAACGGGGC	1076
MF621580.1	TCGAGGGCGCCGTCTGAAGGCTGTCTGATGTCCATCGAGGATGCCGCGAAAATAACGGGGC	1079
MF621576.1	-----GCG	1007
AB915361.1	-----	888

KX250215.1	-----	977
MF621579.1	TCACCGTCGACCAGATCCGCCAGTCGGCGGAATGGATCGCGTCACCAAAGGAAGGCAATG	1067
KT992342.1	TCGAGCAAGCCGAGATCGTCAAGGCAGCCGAGTGGATCGGTATGCCGAAGGAAGGCGGAA	1136
KT992346.1	TAGAGCAAGCCGAGATCGTCAAGGCAGCCGAGTGGATCGGCATGCCGAAGGAAGGCGGAA	1129
KT992341.1	TAGAGCAAGCCGAGATCGTCAAAGCCGCCGAGTGGATCGGCATGCCGAAGAAAGGCGGAA	1136
KT992350.1	TCGAGCAAGCCGAGATCGTCAAAGCCGCCGAGTGGATCGGCATGCCGAAGGAAGGCGGAA	1136
MF621580.1	TCGAGCAAGCCGAGATCGTCAAAGCCGCCGAGTGGATCGGCATGCCGAAGGAAGGCGGAA	1139
MF621576.1	TCTCGGCTTCCGATATCGCGAAGGCGGCACAGTGGATTGCCGAACCCAAGGCCGACGGCA	1067
AB915361.1	-----	888

KX250215.1	-----	977
MF621579.1	CGCGACGCCGCACCATGTTGCCTATGAAAAAGGTATCATCTGGGGCAACGACAAATCGT	1127
KT992342.1	AGCGCCGCCGAGTGATGTTTCGGTTACGAAAAGGGCCTCATCTGGGGCAACGACA-----	1190
KT992346.1	AGCGCCGCCGCGTGATGTTTCGGTTACGAAAAGGCCTGATCTGGGGCAACGATAACTACC	1189
KT992341.1	AGCGCCGCCGCGTGATGTTTCGGTTACGAGAAAGGCCTCATCTGGGGCAACGACA-----	1190
KT992350.1	AGCGCCGCCGCGTCATGTTTGGTTACGAGAAGGGCCTCATCTGGGGCAACGACA-----	1190
MF621580.1	AGCGCCGCCGCGTCATGTTTGGTTACGAAAAGGAATCATCTGGGGCAACGACAAATCGT	1199
MF621576.1	AGCGCCGACGCACCATGCTTCCCTATGAAAAAGGTTTCATCTGGGGCAACGACAAATCGT	1127
AB915361.1	-----	888

KX250215.1	-----	977
MF621579.1	CGAC-----	1131
KT992342.1	-----	1190
KT992346.1	GCACCAACGGCGCCGTGGTGAACCTTGCGCTGGCCACGGGCAATATTGGCCGACCGGGTG	1249
KT992341.1	-----	1190
KT992350.1	-----	1190
MF621580.1	CGAC-----	1203
MF621576.1	CGAC-----	1131
AB915361.1	-----	888

```

KX250215.1      ----- 977
MF621579.1      ----- 1131
KT992342.1      ----- 1190
KT992346.1      GCGGCGTCGTTTCGTTTGGGCGGCCACCAAGAAGGATA 1286
KT992341.1      ----- 1190
KT992350.1      ----- 1190
MF621580.1      ----- 1203
MF621576.1      ----- 1131
AB915361.1      ----- 888

```

c) Xác định các vị trí bảo thủ trên đoạn gen, ghi lại vị trí của các vùng bảo thủ đó với mỗi gen đưa vào so sánh

### Vùng gen bảo thủ ở vị trí nào

Phân tích đa trình tự cho thấy hai vùng bảo tồn nổi bật nằm tại vị trí nucleotide 103–122 và 403–422 trên trình tự gen aioA của *Ancylobacter* sp. TS-1, tương ứng với các đoạn trình tự GTGCACATCGTCATCAAGCC và GGCAAGCTCTATTTTGGGGC.

### 5. Thiết kế mồi bằng Primer-Blast (NCBI)

Sử dụng trình tự của toàn bộ gen quan tâm ở một chủng gần gũi nhất về mặt phân loại với chủng nghiên cứu, đưa vào làm khuôn cho việc thiết kế mồi trên Primer-Blast.

### Lựa chọn các cặp mồi phù hợp và copy kết quả vào ô trống dưới đây

Primer pair 1									
	Sequence (5'→3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	GTGCACATCGTCATCAA	Plus	17	103	119	52.32	47.06	6.00	0.00
Reverse primer	GCCCCAAATAGAGCTTG	Minus	18	422	405	53.34	50.00	4.00	2.00
Product length	320								

6. Phân tích các đặc tính của protein mà gen đó mã hóa, từ đó dự đoán được một phần tính chất của protein trên chủng nghiên cứu. Các đặc tính này bao gồm:

a) Trình tự protein mà gen mã hóa: sử dụng công cụ Translate trên EXPASY

### Kết quả:

DNAFGVDLAEQQGADTPAWYAPSMYNIVRQDGRDVHIVIKPDRECSVNSGLGSVRGARIAEMSYS  
RQRNTQLQRLTDPLVWRYGQLQPTSWDDALDLVARVTSAVIAEQGEDGLFVSAFDHGGAGGGYE

NTWGTGKLYFGAMKVKNIHNRPAYNSEVHGSRDMGVGELNNCYEDAQLADTIVAVGTNALETQ  
TNYFLNHWVPNLRGTSMDKKKAEFGSEPVAQGRVVIVGPRRTVTVNACEVEAGKDNVLHLALNS  
GTDLALFNAWFTYAAEKGWIDREFIGASTKDFDKAVAANKVSIAEAAEITGLSEADIVKAITWIAEP

### Results of translation

- Open reading frames are highlighted in red
- Select your initiator on one of the following frames to retrieve your amino acid sequence

Download all the translated frames

#### 5'3' Frame 1

DNAFGVDLAEQQGADTPAWYAPSMYNIVRQDGRDVHIVIKPDRECSVNSGLGSRGARIAEMSYSRQRNTQLQRLTDPLVWRYGQLQPTSWDDALDI  
VARVTSAVIAEQGEDGLFVSAFDHGGAGGGYENTWGTGKLYFGAMKVKNIHNRPAYNSEVHGSRDMGVGELNNCYEDAQLADTIVAVGTNALETQ  
TNYFLNHWVPNLRGTSMDKKKAEFGSEPVAQGRVVIVGPRRTVTVNACEVEAGKDNVLHLALNSGTDLALFNAWFTYAAEKGWIDREFIGASTKDFD  
KAVAANKVSIAEAAEITGLSEADIVKAITWIAEP

b) Xác định khối lượng phân tử (molecular weight - Mw), điểm đẳng điện (isoelectric point - PI) của protein đó.

Nếu protein sau khi dịch mã có đoạn tín hiệu vận chuyển qua màng, đoạn trình tự này sẽ bị cắt bỏ sau khi được vận chuyển qua màng tế bào, khi xác định khối lượng phân tử và điểm đẳng điện của protein nêu trên, cần sử dụng trình tự protein sau khi đã cắt bỏ trình tự tín hiệu. Tính toán một cách tương đối, khối lượng phân tử của protein có thể được xác định như sau: một amino acid có khối lượng khoảng 110 Da, nếu một protein có 100 amino acid, nó có khối lượng phân tử xấp xỉ là 11 kDa.

Kết quả:

**Theoretical pI/Mw: 5.08 / 35444.57 (Expasy)**

### Compute pI/Mw - Results

#### Theoretical pI/Mw (average)

##### Sequence:

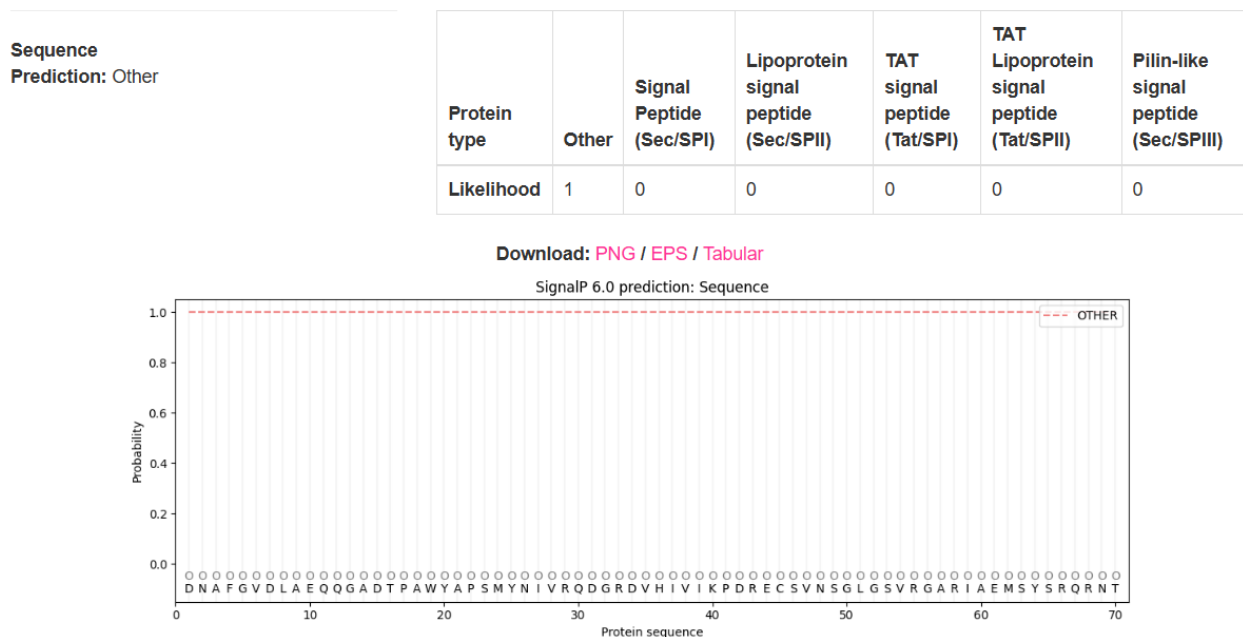
10	20	30	40	50	60
DNAFGVDLAE	QQGADTPAWY	APSMYNIVRQ	DGRDVHIVIK	PDRECSVNSG	LGSVRGARIA
70	80	90	100	110	120
EMSYSRQRNT	QLQRLTDPLV	WRYGQLQPTS	WDDALDLVAR	VTSAVIAEQG	EDGLFVSADF
130	140	150	160	170	180
HGGAGGGYEN	TWGTGKLYFG	AMKVKNIH	NRPAYNSEVH	GSRDMGVGEL	NNCYEDAQLA
190	200	210	220	230	240
DTIVAVGTNA	LETQTNYFLN	HWVPNLRGTS	MDKKKAEFGS	EPVAQGRVVI	VGPRRTVTVN
250	260	270	280	290	300
ACEVEAGKDN	VLHLALNSGT	DLALFNAWFT	YAAEKGWIDR	EFIGASTKDF	DKAVAANKVS
310	320				
IAEAAEITGL	SEADIVKAIT	WIAEP			

Theoretical pI/Mw: 5.08 / 35444.57

c) Quá trình cải biến sau dịch mã (post-translation modification) xảy ra ở một số loại protein và chuyển hóa tiền-protein thành dạng có hoạt tính hoặc vận chuyển protein đến vị trí mà tại đó nó có hoạt tính xúc tác (ví dụ như trên màng tế bào, hoặc trong periplasmic). Một trong các quá trình cải biến protein sau dịch mã là vận chuyển protein qua màng tế bào, thông thường, các protein này có một đoạn trình tự peptide làm tín hiệu (signal peptide), sau khi protein được vận chuyển qua màng ra môi trường hoặc ra periplasmic, đoạn trình tự này bị cắt bỏ. Sử dụng công cụ SignalP 6.0

(<https://services.healthtech.dtu.dk/services/SignalP-6.0/>) để xác định đoạn trình tự protein sau quá trình dịch mã có đoạn trình tự tín hiệu cho việc vận chuyển qua màng tế bào không

**Kết quả: Không**



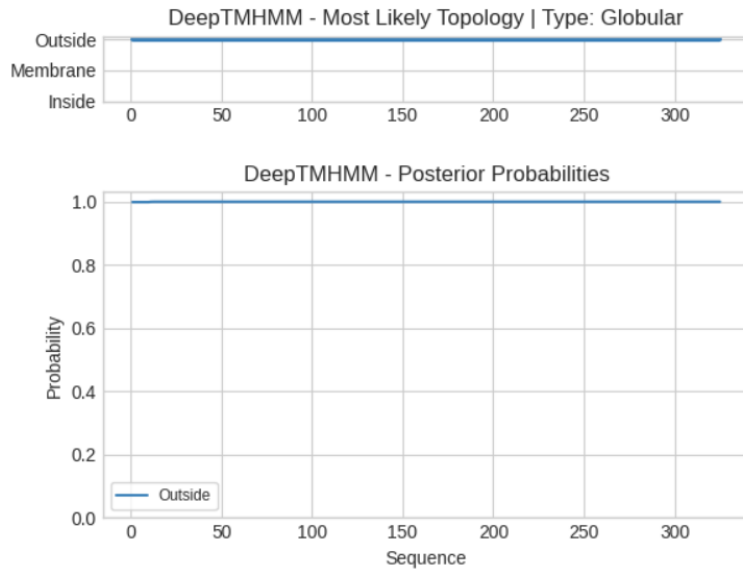
d) Tìm công cụ online DeepTMHMM-1.0

(<https://services.healthtech.dtu.dk/services/DeepTMHMM-1.0/>) để xác định xem protein này có gắn trên màng tế bào (transmembrane) không?

**Kết quả: Không**

## DeepTMHMM - Predictions

Predicted topologies can be downloaded in [.gff3 format](#) and [.3line format](#)



e) Protein được mã hóa bởi gen đó đã được xác định cấu trúc bằng thực nghiệm chưa? Nếu có, bạn hãy tìm cấu trúc trên protein databank hoặc NCBI và tìm hiểu về nó. Nếu chưa, bạn sử dụng Alphafold server để dự đoán cấu trúc của nó.

**Kết quả:**

